DOCUMENT RESUME

ED 025 156 56 EM 006 931

By- Rosen, Marvin J.

An Experimental Design for Comparing the Effects of Instructional Media Programing Procedures: Subjective Vs. Objective Revision Procedures. Final Report.

American Inst. for Research in Behavioral Sciences, Palo Alto, Calif.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No- AIR-F27-5/68-FR
Bureau No- BR-5-8332
Pub Date May 68
Grant- OEG-4-6-058332-1181
Note- 90p.
EDRS Price MF-$0.50 HC-$4.60
Descriptors- Achievement Rating, *Comparative Analysis, Data Analysis, Educational Experiments, Grade 6, Instructional Design, Instructional Media, Lesson Plans, Models, Money Systems, Post Testing, *Program Design, *Program Effectiveness, Programing, Tests of Significance
Identifiers- Jonckheere Procedure

An experiment was designed to compare two alternative procedures for revising a preliminary version of an instructional program, one based on a subjective analysis of the program's behavioral goals, the other on an objective analysis of test data obtained from a trial of the unrevised program. Conforming to an experimental model, 20 programer-instructors were randomly assigned to subjective and objective treatment groups to produce 20 videotaped supplements to the basic prerecorded lesson on English money. The effectiveness of the videotaped revisions was then tested in sixth grade classes, using the mean posttest score in each class as the measure of achievement. The scores produced by the revised programs occurred in the expected order of magnitude (significance factor p .0013), with the original program proving least effective, the subjective revisions more effective, and the objective revisions most effective. No significant correlation occurred between program effectiveness and the amount of either academic training or teaching experience possessed by the programer. The experimental model employed in the study appeared to be a valid guide for assessing the relative effectiveness of programing techniques. A programer's instruction booklet and the script of the basic program are appended to the report. (TI)

BR-5-8332
TR-56
EM006931

Dup.

AIR-F27-5/68-FR

FINAL REPORT
Project No. 5-8332
Grant No. OEG-4-6-058332-1181

AN EXPERIMENTAL DESIGN FOR COMPARING THE EFFECTS OF INSTRUCTIONAL MEDIA PROGRAMING PROCEDURES

ED025156

May 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

AIR-F27-5/68-FR

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

Project No. 5-8332
Grant No. OEG-4-6-058332-1181

AN EXPERIMENTAL DESIGN FOR COMPARING THE EFFECTS OF INSTRUCTIONAL MEDIA PROGRAMING PROCEDURES:

Subjective vs. Objective Revision Procedures

Marvin J. Rosen

American Institutes for Research
in the Behavioral Sciences

Palo Alto, California

ED025156

May 1968

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EM 006 931

TABLE OF CONTENTS

SUMMARY 1

INTRODUCTION 4

Background of the study
An experimental model for comparing programing procedures
An application to compare program revision procedures
Review of related research
Objectives of the study

METHODS AND PROCEDURES 17

General design
Definition of "revision"
Selection of media
Selection of grade level
Development of the experimental "base" program
Development of post-test
Standardization of treatments
Administration of program trials

ANALYSIS AND FINDINGS 40

Problems of analysis
Jonckheere's procedure
The significance test for an obtained value of S
Analysis of data: H_0 vs. H_1
Analysis of data: H_0 vs. H_2 and H_3
Summary of findings

CONCLUSIONS AND RECOMMENDATIONS 50

REFERENCES 54

APPENDIXES

A. Thirty-item test used for item analysis 57
B. Instruction booklet: Introductions, program objectives, test-item "generator", script of base program, achievement test. 65

LIST OF TABLES

1. Operational definitions of S-R elements as used in the base program 24

2. Test-item generating statements 26

3. Summary of class means from program trials by treatment 43

4. Summary of analysis of ordered differences among sample means 44

5. Summary of correlations (r_s) between program effects and selected programer variables 48

6. Summary of ranked data used for calculation of r_s. 49

LIST OF FIGURES

1. Diagram of experimental model for comparing program characteristics 5

2. Diagram of experimental model for comparing programing procedures. 7

3. General design of the experiment 18

4. Photograph of standard studio arrangement 30

5. Photograph of standard program planning arrangement 31

6--25. Photographs of selected scenes from the revised programs 35ff

26. Cumulative distribution of class means by treatment 45

ACKNOWLEDGEMENTS

The hundreds of sixth graders who so graciously submitted to our antics deserve to be publicly thanked, both for their patience during the program and for their scrupulously honest, yet ever so gentle criticisms following. They should never be underestimated.

Their many teachers, too, and school administrators who so willingly squeezed out precious time from both personal and school schedules to make this work possible deserve more than I can repay.

Particularly, I wish to acknowledge the very large contributions of Arthur A. Lumsdaine, under whose tutelage this project originally took form, and Leslie J. Briggs, G. Kasten Tallmadge and Theodore R. Husek all of whom, in their own ways, kept me going.

Marvin J. Rosen
Project Director

SUMMARY

An important goal of research on teaching is the generation of principles and procedures which result in the improvement of instruction. It is one thing to state a theoretical principle of learning or a characteristic of programs which can be shown to operate in effective instructional sequences; it is quite another for teachers, or programers, to translate these statements into procedures for developing more effective instructional programs. Yet the usefulness of teaching principles lies exactly in their power to improve the effectiveness of instruction.

The design of most experiments which compare program characteristics and principles provide no basis for assessing if programers, other than the experimenter, can incorporate the characteristics into their programs with good effects. A design which would provide for such an assessment should show that a number of programers following some prescribed procedure can independently produce consistently superior programs. By providing the means for assessing the variance obtaining from programer differences, a body of procedural techniques for effective programing might readily evolve.

The object of this study was to conduct a pilot experiment applying an experimental model which identified programers as the experimental units of interest, the procedures given them to follow (instructions) as the independent variable, and the effectiveness of their products as the dependent variable. The substantive question raised in the study concerned whether a procedure for revising an instructional program could be developed which would lead a sample of programers to produce revised programs which were reliably more effective than (1) the original program before revision, and (2) revisions produced by a sample of programers following some reasonable alternative procedure.

The experiment consisted of an after-only comparison of the effectiveness of programs prepared by two independent samples of programers following two sets of procedural instructions. The two procedures compared may be described as a "subjective" vs. an "objective" procedure for analyzing a program as a basis for revising it. The relationship of academic preparation and prior teaching experience to the programer's effectiveness were further questions of interest in the study.

A 25-minute slide-tape program on the subject of English Money was developed to serve as the base program. Two statements of the program's objectives were prepared, one of which provided general statements which could be used to generate a large population of test items, and the other of which provided a specific sample of test items. The base program was administered to a sample of sixth grade students together with the sample test items. The data obtained were

analyzed for difficulty and discrimination power to select the items which were used as the post-test of the program.

Twenty teacher-subjects (S's) were then recruited, each of whom was to prepare a 15-minute videotaped supplement for use following the base program. The S's were randomly assigned to two treatment groups and each directed to appear for a day at the television studio to plan and produce the lesson. The conditions under which the lessons were prepared were standardized for all S's: no S was informed of the subject matter until arrival for production; all S's were limited to a maximum of six hours production time from start to finished tape; all work was performed in the studio from the materials provided by the experimenter; only technical assistance was provided by the television staff (directorial assistance being limited only to correcting gross errors in content); and tapes were limited to 15-minutes in length as an absolute maximum.

The instructions given to Treatment Group I (subjective) directed the programers to study the general objectives of the base program as characterized by the item-generating statements, then to view the base program estimating from their own teaching experience what additional instruction was most needed to enhance the effect of the total lesson. Based upon this estimate, then, they were to plan a 15-minute supplementary lesson to be recorded on videotape which would subsequently be used with the base program to teach a sample of sixth graders.

The instructions given to Treatment Group II (objective) were essentially the same except that the final test items were provided together with data showing the percentage of sixth graders who actually succeeded on each item after a trial run of the base program.

Twenty videotaped supplements were thus prepared and subsequently randomly assigned to 20 sixth-grade classes for administration. Each administration involved presenting a standard introduction, the base program, one of the twenty videotaped supplements, and a standard post-test under time-limited conditions.

The effectiveness score assigned to each program was based upon the mean post-test score of the class in which the program was shown. As each program was shown in only one randomly assigned class, N = 11 for Treatment Group I, N = 9 for Treatment Group II, and N = 7 for the Base Program Only Group.

To test the null hypothesis against the alternative that the magnitude of the effects produced by the different treatment groups occurred in a predicted order, a special application of the statistic S was employed. The scores obtained from all the treatment groups revealed that the revisions based upon an analysis of behavioral objectives were significantly more effective than the original version of the program. Moreover, the revisions based upon an objective analysis of test data obtained from a trial run of the original program were significantly more effective than the revisions based solely upon a subjective analysis of the objectives.

An analysis of the scores in relation to the amount of the programers' academic preparation and teaching experience failed to reveal any statistically significant correlation.

It was concluded that (1) the two revision procedures employed in the experiment led programers to produce revised programs which were more effective than the original, (2) of the two procedures employed, the one which provided empirical test data from a tryout of the original program led to the production of more effective revisions than the procedure which provided only general statements of the program's objectives, (3) the experimental model employed in the study appears to be a viable procedure for assessing the relative effectiveness of alternative programing techniques.

INTRODUCTION

Background of the Study

General -- An important goal of research on teaching is the generation of principles and procedures which result in the improvement of instruction. The findings of teaching research should provide to persons who produce teaching materials -- teachers, film makers, script writers, editors, hereinafter termed "programers" -- with viable procedures by which their products, termed "programs," might reliably be improved. If the principles which evolve from teaching research cannot be demonstrated to improve the effectiveness of programer performance, the usefulness of our research efforts may be severely limited.

It is one thing to state a theoretical principle of learning, or a characteristic of programs, which can be shown to operate in effective programs; it is quite another for programers to transform these statements into procedures by which they can consistently improve their products. Yet the usefulness of such statements lies exactly in their power to be translated into procedures which can improve the programer's effectiveness. There is little evidence, however, that the principles of learning and the characteristics of programs being generated by our research efforts can, in fact, be used with good effects by others.

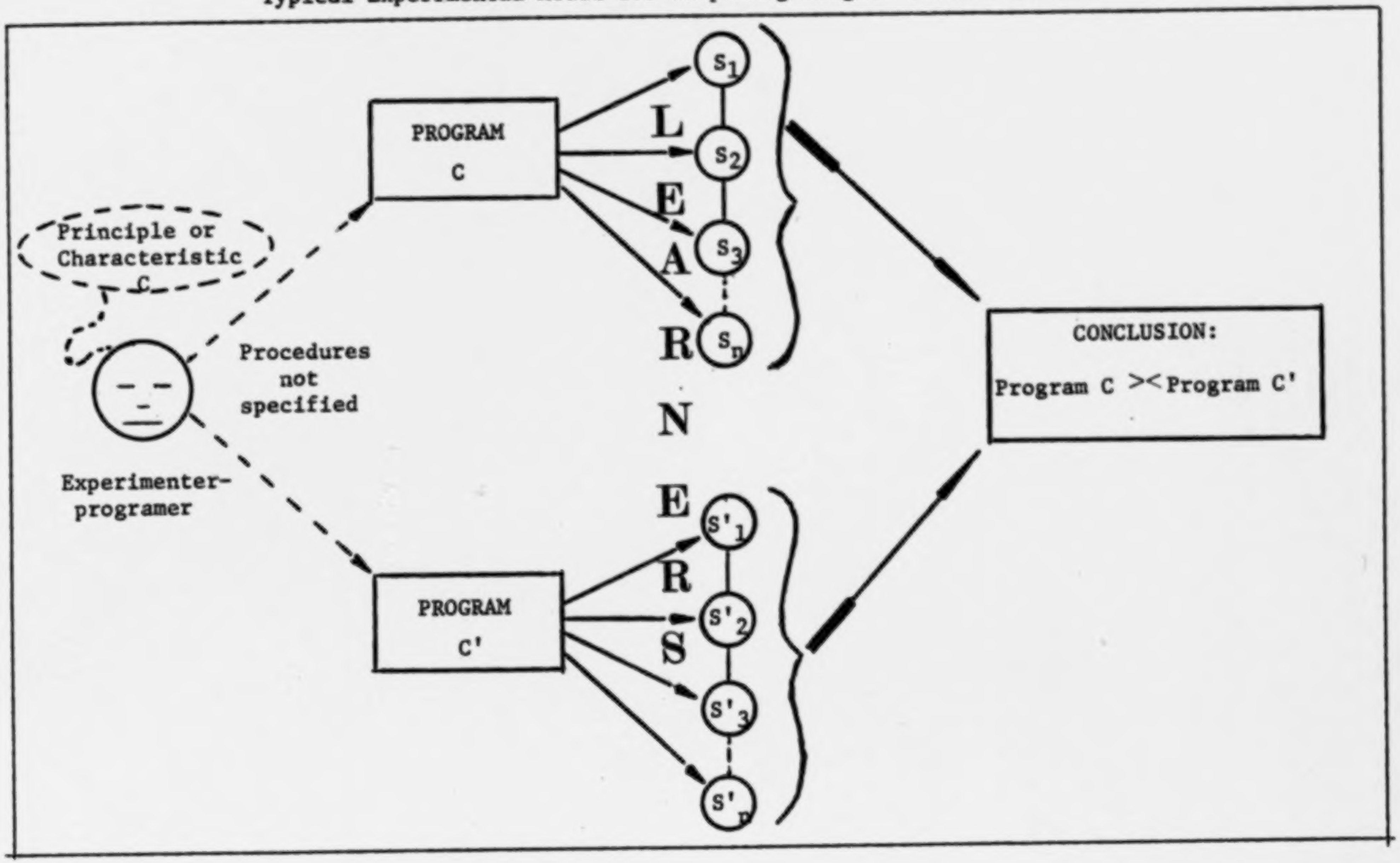
Most of the research relevant to instructional program variables has come out of the instructional media fields. The instructional media, particularly self-instructional programs afford excellent tools for the systematic application of learning and programing principles. By standardizing stimulus sequences, the experimenter is provided greater control over certain classes of instructional variables than is usually feasible in the classroom.

For the most part, experiments employing these media have compared two or more versions of a program in which some program characteristic has been systematically varied by the programer. The characteristic varied is usually described as an implementation of some learning principle. In this way, the researcher hopes to generate evidence supporting or refuting the validity of the principle (see Figure 1).

A popular experiment with instructional programs, for example, compares the effects of variations in "active response." Postulating that students who respond actively more frequently during instruction learn better than those who respond actively less frequently, the experimenter develops two programs which presumably differ only with respect to the amount of "active responding" they occasion. The programs are then administered to students and measures of their learning are obtained and compared. If the students are found to differ on the dependent measure, the difference is

FIGURE 1

Typical Experimental Model for Comparing Program Characteristics


Principle or
Characteristic
C
Experimenter-
programer
Procedures
not
specified
PROGRAM
C
PROGRAM
C'
LEARNERS
S_1
S_2
S_3
S_n
S'_1
S'_2
S'_3
S'_n
CONCLUSION:
Program C >< Program C'

attributed to the different amounts of "active response" engaged in by the learners (see Lumsdaine, 13, pp. 609-15).

The conclusion to which we are led, of course, is that programs which occasion more active response mediate learning better than those which occasion less active response. However, the experimental model employed above allows little formal basis for such a generalization. Any number of alternative hypotheses might be offered to explain the difference between any two programs, e.g. the programer "felt better" writing one than the other, he wrote a biased program favoring the desired outcome, one program contained additional or different stimulus materials which produced variations other than "active response," and so on. In point of fact, an experiment which compares only two programs provides no basis upon which to reject the null hypothesis that one program was superior to the other simply by chance.

An Experimental Model for Comparing Programing Procedures -- Generally, we hypothesize that implementing some instructional principle will have a beneficial effect upon learning, i.e. can lead programers to increase the effectiveness of their products. Therefore, we should seek to generalize the viability of the principle across some population of programers and their products.

To this end our experiments should provide some formal basis for generalizing to this population. If some instructional principle is to be shown to have beneficial effects upon learning, it must be shown that procedures for its employment can be communicated to independent samples of programers with consistent good effects. It must be shown that a number of programers following these procedures can independently produce programs which are consistently superior to those produced by other programers following alternative procedures. Only then can the variance obtaining from individual programer differences be formally assessed.

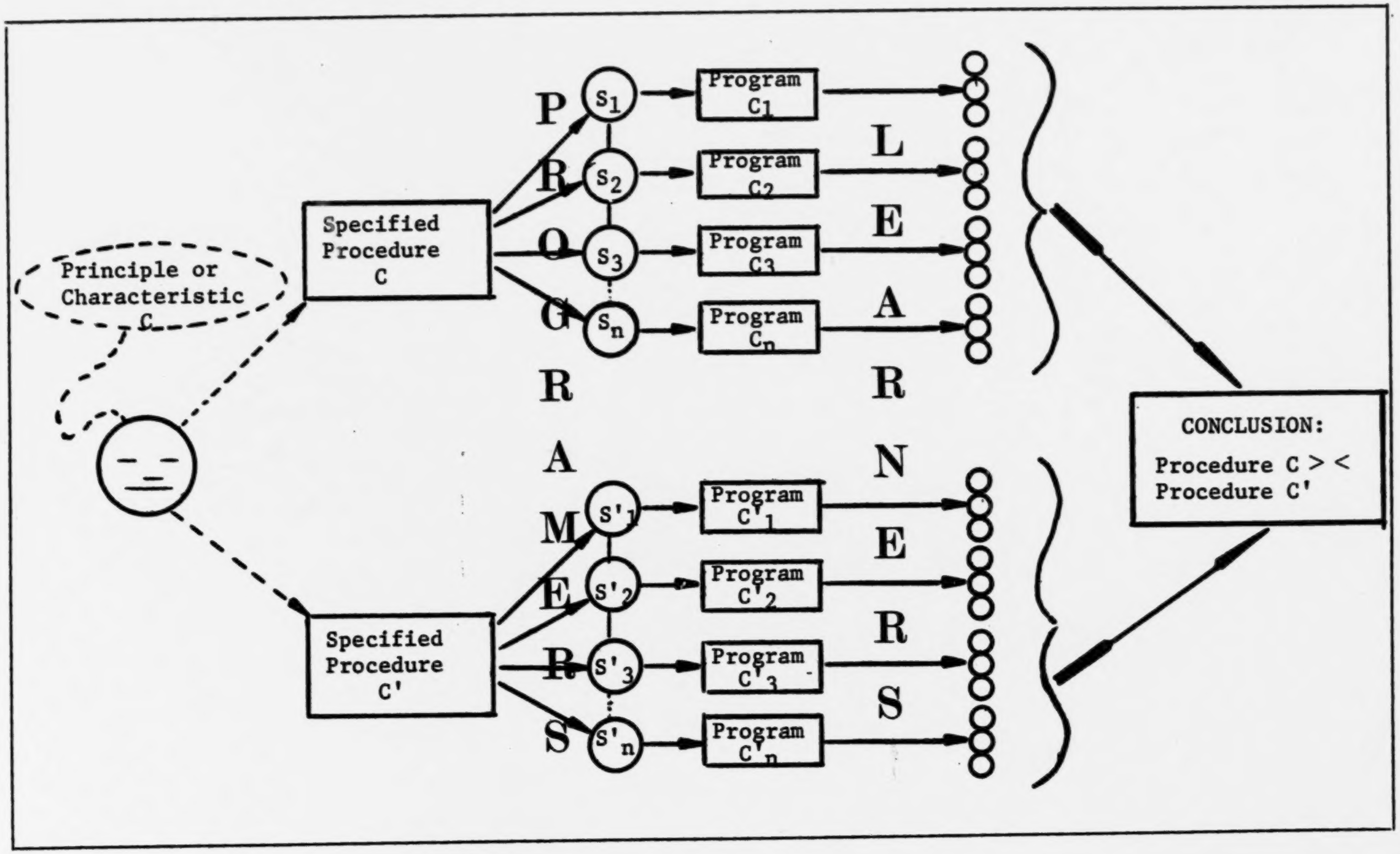
This suggests an experimental model in which programers are the experimental units, the effectiveness of their products is the dependent variable, and the procedural instructions for implementing the programing principle constitute the experimental variable of interest (see Figure 2).(See Lumsdaine,14 ,p. 386.)

It was the purpose of this study to employ such a model in a trial experiment assessing the effectiveness of certain programing procedures. If it could be shown that the instructions given to a sample of programers could materially improve their products, future programing research might profitably be directed toward comparative studies of alternative programing procedures.

An Application to Compare Program Revision Procedures -- In the application of this experimental model to a specific problem, it

FIGURE 2

Experimental Model Employed to Compare Programing Procedures


Principle or
Characteristic
C
Specified Procedure C
Specified Procedure C'
PROGRAMERS
S_1
S_2
S_3
S_n
S'_1
S'_2
S'_3
S'_n
Program C_1
Program C_2
Program C_3
Program C_n
Program C'_1
Program C'_2
Program C'_3
Program C'_n
LEARNERS
CONCLUSION:
Procedure C > <
Procedure C'

was proposed to compare two procedures for revising an audio-visual instructional program to improve its effectiveness.

Traditionally, developmental revision of instructional materials has been based upon the subjective analysis of authoritative persons. The revision of pre-publication drafts of instructional materials by subjective analysis, for example, has been common practice at many levels of program development activity. At the commercial level, for example, editors and publishers of textbooks may be found generally to follow this traditional procedure by employing "experts" to consult with them upon the effectiveness of the materials at several points in their development. These consultations form the basis for corrective modifications of estimated weaknesses in the materials. At the classroom level, on the other hand, teachers will be found to "preview" educational films and to base their supplementary instruction upon subjective estimates of the film's effectiveness.

This type of subjective procedure has been challenged by advocates of more empirical analysis procedures (see Rothkopf, 23; and AERA, 1). Given a clear statement of the effects a program is intended to produce, they say, the more appropriate measure of its effectiveness is an analysis of the program's observed effects upon the behavior of students. They advance the notion that the program's objectives ought to be stated so as to imply clearly a test of learned performance and that the analysis of student responses in the test situation is an effective basis upon which to correctively modify the program (see Gropper, et al, 4).

While the principal issue of this study lies in the application of a particular experimental model for comparing the effects of alternative programing procedures, the selection of a problem in the area of analysis-for-revision may be seen to have important implications for a wide range of programers from the classroom teacher to the mass-media programer. In the creation of instructional materials, judgements concerning their effectiveness are made more or less continuously upon which the materials are subsequently modified. In this sense the task of the programer seeking to revise an early draft of a program is little different from the task of the teacher seeking to supplement some instructional material for her students. Each assumes that the product in hand is insufficient for producing all the desired effects and each seeks to modify the material on the basis of an estimate of the nature of that insufficiency. Each has in mind some set of objectives, either expressed or implied, and each acts to modify the material to better achieve these objectives.

In view of the great emphasis self-instructional programers place upon the analysis of student response data as a means for revising programs, empirical evidence might be thought to abound substantiating the effectiveness of such procedures. In spite of the apparent number of programs that have been produced by such methods, however, there is little experimental evidence that the final versions of the programs so produced were in fact more effective than the initial versions. Further, even granting that the final versions were superior, there is virtually no evidence to support the claim that these procedures produced final versions superior to those which might have been produced by a more traditional and less costly procedure. Finally, and with particular regard to the principal issue of this study, there is no evidence that any specific procedure for revising a program can be communicated to others such that the programers who follow it may reliably be expected to improve upon the original.

The proposed study may be described as a comparison of two procedures, one "subjective" and one "objective" for improving programer effectiveness in revising an instructional program.

Review of Related Research

A review of the experimental literature dealing with procedures for analyzing programs as a basis for revision was conducted to provide guidance in designing the experimental materials and procedures which would be used in the study.

Audiovisual programs have been modified on the basis of analysis of student response data by a number of researchers who report upon the effectiveness of their procedures. Most of these studies have been conducted to effect two kinds of outcomes upon an intended audience: (1) attitude change and, (2) cognitive gain. Most of the formal effort has been expended in the former.

Sheffield (24) conducted extensive prior interviewing among an intended soldier population to determine the bases of resistance to attitude change. His analysis of the soldiers' responses to the interview questions was employed in the preparation of an audio recording intended to modify soldier attitudes about the probable length of the war.

The procedure involved, among other "rules," an attempt to introduce negative arguments at those points where they seemed likely to occur spontaneously to the opposed members of the audience. These points were determined empirically by pretesting a sample of the expected audience by means of individual interview techniques. In these pretests, interviewers spoke individually to the soldiers and actually presented the case to be advocated by the subsequent transcription. They attempted to elicit counter-arguments from interviewees who felt the war would be short. As much as possible the points at which the counter-arguments were raised were the points used in the transcription for raising the same counter-arguments. The outline of factual materials organized in this way was thus used by the script writers as a basis for preparing the program that used arguments on "both sides" of the question. It should be noted that, at the time of preparing the scripts, the writer knew the purpose of the experiment and the actual wording of the main question to be used in measuring the effects of the transcription.

Among men initially opposed to the position advocated by the transcription, the "one side" message produced a net change in the direction advocated by the message of 36%. The "two side" message (which had been prepared after prior interviewing of the intended audience) produced a net change in the desired direction of 48%.

In this case then, prior interviewing of an intended audience was used to determine the nature of opposing attitudes and the stimulus conditions which aroused them. Theoretical considerations were then employed to determine an appropriate use for this data in constructing an attitude-changing message.

More recently, Levonian (10) conducted extensive prior testing of the opinion dimensions toward India of an intended film audience. Prior to scripting the film he derived a number of relatively independent opinion dimensions toward India by factor analyzing responses to a 69-item opinion questionnaire. Each item in the questionnaire (itself a film) consisted of a film scene, a narrated question, and an abbreviated repetition of the narrated question printed on a mimeographed form in the student's hand. Each scene lasted about 15 seconds, during which time the narrated question was asked. At the end of each scene a title, ANSWER ITEM ___, appeared on the screen in black letters. The projection time of the title allowed the respondents about 15 seconds to answer each item.

A factor analysis of the responses revealed ten common opinion dimensions, of which eight appeared related to specific content areas. These eight opinion dimensions formed the basis for the subsequent development of the film. Before constructing the film, the researchers decided upon the orientation to be adopted for each content area. This orientation was coincident with the direction of the desired change of opinion.

To test the effects of the final film, Levonian (11) selected 540 university students, each of whom was given a 36-item questionnaire pertaining to the material presented in the film. One experimental group was shown the film a week later and the questionnaire was readministered immediately after the film. The other experimental group was also shown the film a week later, but the questionnaire was readministered one week after the film showing. The questionnaire was readministered a week later to the control group, which did not see the film at all. Relative to the control group, the experimental groups showed large changes of opinion in the desired direction.

Edling (3) conducted an experiment to determine if films produced with a knowledge of the learner's motives were more effective in changing attitudes among those for whom the materials were intended than among individuals in general. Motives were operationally defined in terms of responses to existing psychological instruments. A target group of high ability high school students who indicated no plans for post-high school education were tested on these psychological instruments and their responses subjected to a factor analysis. Large differences in learner responses by sex, plus the dictates of the research design, led to the production of two films, each intended to modify attitudes toward post-high school education for a group with specific motivational characteristics. It was demonstrated that the film intended for the student group with a specific motivational pattern mediated greater opinion change for that group than for people in general or for groups possessing other motivational pattern characteristics. Among other things, he concluded that techniques employed by the behavioral scientists can make a significant contribution to the development of more effective instructional materials at two points: first, in the analysis of learner characteristics as a basis for structuring content and, second, in empirically testing materials to determine their psychological validity.

In addition to these attitude-changing programs, the analysis of student responses to content-related questions has been employed to develop programs intended to produce increments of cognitive gain. Procedures of this sort are becoming more common, especially among developers of self-instructional programs. To a limited extent, extensive analysis of data both from pretests and from within-program responses is being used as a basis for revision by some program developers. (See AERA, 1, Lumsdaine, 14, and Markle, 17.)

J. V. Zuckerman (30) pointed out as early as 1951 a high correlation (Pearson r's of .71-.94) between post-test scores obtained by one group of subjects which saw a finished color version of a film, and the post-test scores of a second group which saw a rough, black-and-white filmstrip of the storyboard of the same film.

The study was concerned with ascertaining experimentally the extent to which the responses of typical students to a preliminary filmstrip, made from a series of "storyboard" sketches, could serve as a predictive device to tell how well students would learn specific points of information from a completed film. The sketches were intended to represent the most important visual aspects of the scenes to be photographed in the motion picture, "Flight Capabilities of the F86-A," with the accompanying dialogue or narration indicated for each sketch. Two groups of subjects were compared: Group II viewed the completed color motion picture. The same sound track was used for both groups. Both groups were given pre- and post-tests covering in detail the points of information presented in the films. Subjects were 90 Air Force student pilots in advanced single-engine pilot training. The results, based on percentages of the correct answers for each of the 26 multiple-choice test items, were computed in terms of several adjusted scores. The level of learning obtained varied widely among the various points of information covered by the films.

Pearson product-moment correlation coefficients, computed between the pairs of item percentages for the two treatments, were .80, .71, .94, and .89, respectively, for four measures.

Regardless of which measures are considered, the varying levels of knowledge on specific points obtained from the film are thus predicted fairly well by the levels of knowledge obtained from the filmstrip. This means that tryout data from the preliminary form could have been used profitably for identifying how well various points of information would get across and, hence, as a basis for knowing on which points it should be strengthened. An incidental finding was that the average level of achievement for the final, polished film was roughly the same as for the crude preliminary version. (See also May and Lumsdaine, 19.)

Gropper, Lumsdaine and Shipman (4) made application of this finding to demonstrate how such a rough, preview version of a lesson might be used as a basis for gathering student response data upon which to revise the lesson prior to producing the final version.

Two different lessons - one on the effects of heat and one on an introduction to chemistry - in a junior high school science series televised by the Metropolitan Pittsburgh Educational Television Stations were shown on separate occasions to a preview audience of students. Following the preview showing, students took an achievement test based upon the lessons. Analysis of the items revealed which points were misunderstood or poorly learned. The lessons were then analyzed to discover the possible reasons for their ineffectiveness, and were revised to remedy the weaknesses. Later, both the preview version and the revised version were telecast simultaneously on the day the lesson was regularly scheduled to be shown. Students who watched each version then took an identical achievement test.

Results of the experiment showed an average improvement varying from 12 to 26 percent for students who watched the revised versions over students who watched the original preview versions. It was concluded that,as feedback to the programer, test results following a preview of a lesson are an effective substitute for the interaction between student and teacher that occurs in a classroom. The procedure enables the programer to tailor his presentation to the needs and abilities of his particular student audience.

VanderMeer and his associates (28 and 29) report upon a similar procedure for using post-test data derived from early versions of both filmstrips and films to produce revisions more effective than the original. In one case they developed tests based upon the content of two existing films and subsequently revised the films on the basis of student responses on the tests by adding additional instruction and modifying both the graphic and aural content of the films. Two successive revisions were made, the second of which produced significantly superior performance over the original version. In the other case, an existing filmstrip was revised by a similar procedure with similar results.

Markle (18) reports using student performance data to support many design decisions in the development of a multi-media basic first aid course intended, in seven and one-half hours, to produce results at least equivalent to those produced by a standard conventional ten-hour course then in current use. The course objectives were developed by pretesting a sample of the target population to determine levels of entering knowledge. A revised set of test questions was then used as a first draft of the course, instructional elements being added on each topic until trial students could respond successfully to the questions. Each instructional element (film, filmstrip, worksheet, or other) was pretested with trial students and revised in response to the test results prior to final publication.

The mean scores of students taking the resulting course were as much as 86% higher than those of students who took the older standard course, and the time spent in the program had been reduced by 25%. In the process Markle produced three successive developmental versions of the course, the first was a preliminary trial version, and each of the succeeding versions was a revision of the one preceding it. The methods employed, using student responses to test questions based upon the content, showed each succeeding version to be more effective than its predecessor in less instructional time.

Lumsdaine and May (15, p.494) point out that surprisingly few studies have been reported to demonstrate empirically that the expected value of successive revision based upon feedback of test data to the programer has actually been realized. Earlier versions of a revised program are generally not preserved after the revised version has been completed. Though the revised version is generally felt to be an improvement over its earlier counterpart, this fact has rarely been demonstrated. While both Edling (3) and Levonian (10) felt that films "tailored" to student characteristics were more effective in modifying attitudes than films not so tailored, neither made a direct experimental comparison between two such films. In another study Holland (6) compared student errors within a program and time to completion for both a revised and original version of a self-instructional program, but did not report a comparison of post-test data derived from students in each of the programs.

The first such demonstration appears to have been by Gropper, Lumsdaine and Shipman (4) who compared simultaneous telecasts of earlier and later revisions of a video-taped program and showed appreciable gains from both a first and second program revision based upon the results of achievement tests. (See also Gropper and Lumsdaine, 5.) In later studies, both Silberman (26) and Markle (18) replicated these findings in the development, in one case, of self-instructional programs and, in the other case, a multi-media course of instruction.

Objectives of the Study

Three important questions emerge from these lines of inquiry. First, given that the framing of items for a test of a program's content may be regarded as a process of specifying behavioral objectives, we might ask whether an experienced teacher given a clear specification of objectives might not predict, *a priori*, those items on which students will likely fail. None of the studies cited above compared their objective procedures with such an alternative subjective procedure.

In this regard there is some evidence to suggest that teachers might not succeed in this kind of analysis, in spite of the popular assumption to the contrary. Rothkopf (23) conducted an experiment to determine the ability of "experts" to predict the effectiveness of an instructional program by reading it. The actual effectiveness of seven versions of a program were previously determined experimentally. Twelve educators who had completed a course on programing were then asked to read the seven versions and rank them according to their predicted effectiveness. The predictions of these twelve authoritative persons correlated -.75 (sic) with actual effectiveness scores. In other words, a reasonably good prediction of effectiveness could be obtained by predicting the reverse of these expert predictions (see also Merrill and McAshan, 20).

Hovland, et al, (7) relate similar experiences when trying to obtain constructive suggestions from viewers on how to improve a film's effectiveness. While readily expressing liking or disliking an educational film, they point out, the viewers were singularly inarticulate as to their reasons and had few suggestions of ways to improve it. Similar findings were reported by Twyford (27) in an experiment to determine if a subjective rating profile could be considered a reliable measure of learning from a film. None of his self-report measures could be used reliably to predict actual learning mediated by the film.

Still the question is an important one from the point of view of the cost benefits to be derived from exhaustive testing of preliminary versions of an instructional program to produce a more effective final product. One of the objectives of the study, therefore, was to compare experimentally the relative effectiveness of two procedures for revising an instructional program, one based upon a subjective review of the program with respect to a set of behavioral objectives and the other based upon objective test data obtained from an actual tryout of the program.

A second question to emerge from these studies concerns whether certain psychological or demographic characteristics are related to the ability to act effectively to improve a program's effectiveness. It is clear from most of the studies that the experimenters applied objective procedures and produced effective programs. The relationship between these two outcomes, however, does not appear to have been assessed. It is not clear, for example, if their good results were due to their procedures, per se, or to some personal characteristics which enabled them to interact with the data in some unique way. Knowing what the trouble is, even knowing where it is, does not ensure that anyone can correct it. None of the studies cited formally compared programers on their ability to act effectively upon student response data (see Meyers, 21).

Another objective of the study, therefore, was to relate two of the demographic characteristics most commonly felt to affect teaching ability -- academic preparation and teaching experience -- to the effectiveness of the programer's effort. To some extent it should be expected that programers with more education and experience in teaching would have a superior basis upon which to prepare a supplementary revision than the less experienced programer. The a priori analysis of an experienced teacher should predict the actual performance of students better than the analysis of a less experienced teacher. At least that is the popular notion. Further, given actual test data pinpointing specific weaknesses in the program, it should be expected that the more experienced programer might act more effectively than his less experienced counterpart to overcome those weaknesses. In other words, the more experienced teachers should be expected to revise programs more effectively than the less experienced teachers irrespective of the procedures employed.

Finally, the most important issue arising from the research and the major concern of this study concerned whether a procedure for revising a program could be specified which would lead a sample of programers to consistently superior results. In the studies cited above it was demonstrated only that the experimenters produced effective programs. In none of the studies was it formally demonstrated that other programers following the same procedures would be likely to produce the same effects. Nor was it demonstrated that the use of objective student response data was instrumental in obtaining the improved effects (see Lumsdaine, 16, p. 292). The major objective of the study, therefore, was to compare the effectiveness of two alternative programing procedures as they would be employed by a sample of programers to revise an early version of an instructional program.

These questions were expressed in terms of the following experimental hypotheses which were the object of the study:

H_1: Given a preliminary version of an instructional program and a set of its behavioral objectives, programers will be able to revise it effectively on the basis of a subjective analysis; moreover, those who additionally study tryout test data will produce more effective revisions than those who do not (original version < subjective revision < objective revision).

H_2: Amount of teaching experience is positively correlated with the effectiveness of the programers' revisions (high experience > low experience).

H_3: Amount of academic preparation is positively correlated with the effectiveness of the programers' revisions (high education > low education).

METHODS AND PROCEDURES

General Design

An experiment was designed to compare the effects of two alternative procedures for improving a preliminary version of an instructional program, one of which provided only general statements of the learning objectives and the other of which provided achievement test data obtained after a trial run of the program.

Learning tasks were analyzed and a set of behavioral objectives were established for a fixed-pace instructional program consisting of a set of slides coordinated with a tape recording on the subject of English Money for use in sixth-grade classes. An achievement test was developed based upon the stated objectives; then trials of the original program were conducted using the test to obtain after-only measures of the program's initial effectiveness.

Twenty programers, selected from a population of state-credentialed teachers, were randomly assigned to one of two treatment groups: I (subjective) in which only general statements of objectives were provided and II (objective) in which achievement test data obtained from the trial run was provided. Each programer then planned and videotaped a 15-minute supplementary lesson based upon his analysis of the information provided.

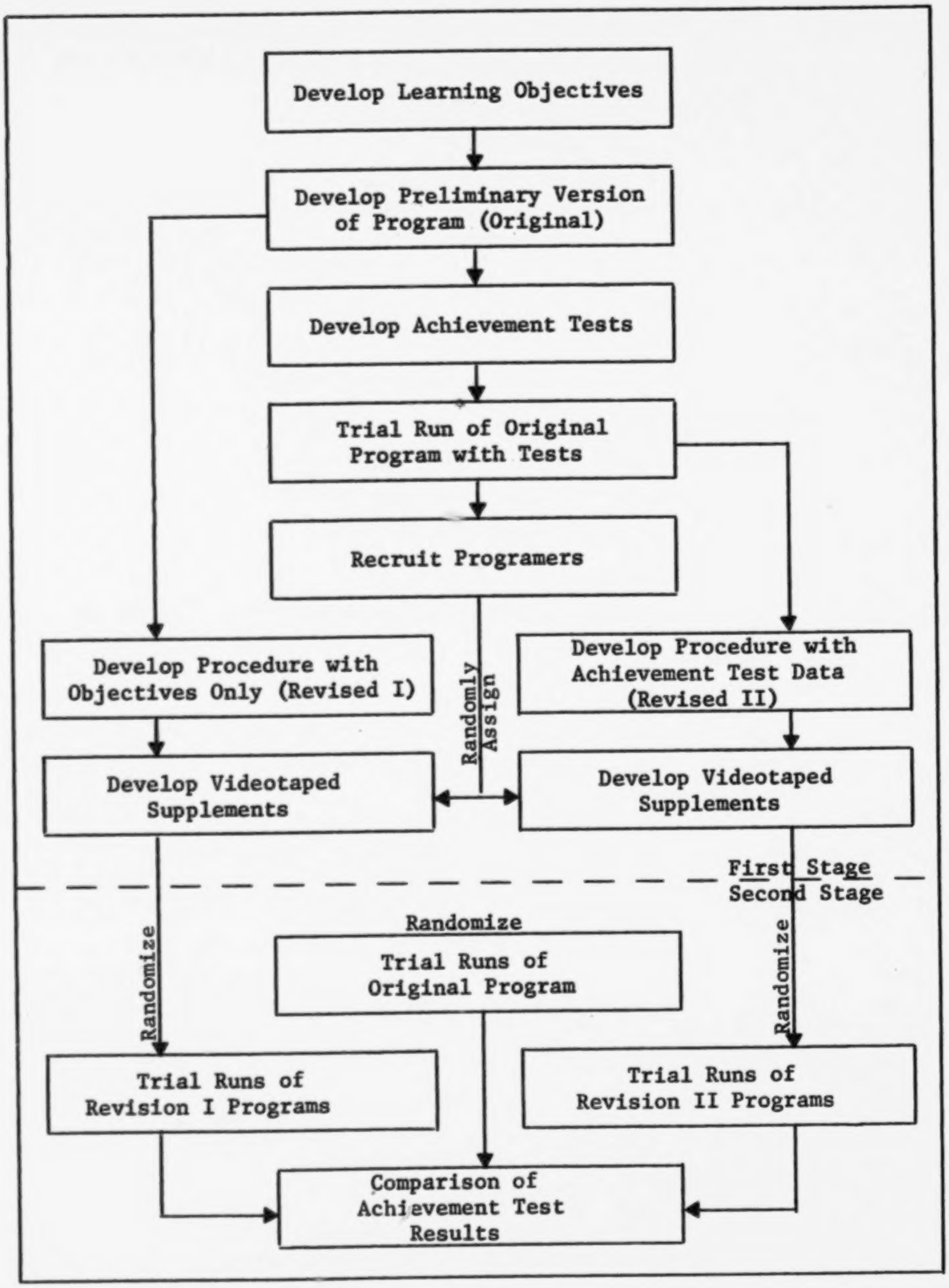
Trials of the 20 resulting videotapes were conducted in 20 randomly assigned sixth-grade classes, following which the same achievement test was given to obtain measures of effectiveness for each lesson. The effectiveness of each videotaped lesson was represented by the class mean of the scores obtained.

Entire classes, rather than individuals, were assigned at random for trials of the videotaped lessons since it was not administratively feasible within the framework of the public schools of the area to use individual sampling procedures. Since entire, intact classes were sampled, the class rather than the individual student became the proper unit for observation and analysis. Thus, the data associated with each program trial were statistically treated using the class rather than the student as the unit of observation (see Lindquist, 12) in order to determine the relative effectiveness of the treatment groups.

The general method of the experiment is diagramed in Figure 3, showing the major flow of activities involved in the study. Generally, these activities involved two stages: (1) the development of program and test materials within the framework of experimentally controlled conditions and (2) the trials of the experimentally produced materials to obtain measures of their relative effectiveness.

FIGURE 3

General Design of the Experiment


Develop Learning Objectives
Develop Preliminary Version of Program (Original)
Develop Achievement Tests
Trial Run of Original Program with Tests
Recruit Programers
Develop Procedure with Objectives Only (Revised I)
Randomly Assign
Develop Procedure with Achievement Test Data (Revised II)
Develop Videotaped Supplements
Develop Videotaped Supplements
First Stage
Second Stage
Randomize
Randomize
Randomize
Trial Runs of Original Program
Trial Runs of Revision I Programs
Trial Runs of Revision II Programs
Comparison of Achievement Test Results

Thus, the general design of the experiment consisted of an after-only comparison of the effectiveness of programs prepared by two independent samples of programers using different procedures. The experimental units of interest were the products of the programers and the independent variable was test data provided to the programers in the set of instructions given for them to follow. The dependent variable was a post-test measure of the effectiveness of their programs. In addition, the amounts of education and teaching experience were correlated with program effectiveness.

Definition of "Revision"

At an early stage a determination of the exact nature of the "revised version" had to be made. For procedural reasons in this study, "revision" was defined as "the addition of supplementary elements of instruction to a preliminary version of a program." This definition was selected rather than one which would permit deletions and rearrangements of the base program to simplify both the production and presentation of the revised versions. While this constraint imposed a somewhat artificial condition upon the work of the programer, who developmentally revises the internal structure of a program, the condition was quite analogous to the kind of revision practiced by the classroom teacher who is constantly asked to provide supplementary instruction to existing published materials. It did not seem, therefore, that this constraint would limit the generalizability of the findings since it was a revision strategy under consideration, not a program form. If one strategy was shown to lead programers to superior revision decisions, we would expect this to hold true regardless of the forms of the programs being revised. On the other hand, this constraint reduced considerably the complexity and thereby the cost of producing the revised versions.

The revised version therefore was defined as consisting of two elements: (1) the original experimental base program and (2) a supplementary sequence of instruction designed to follow the base program.

Selection of Media

The most common revision situation is that which involves supplementing an established sequence of instruction, such as a film or published book. The choice in this study was to simulate an established sequence presented on a fixed-pace basis. The choice of a fixed-pace sequence was governed partly by the anticipated need to administer the program under time-limited conditions in schools, and partly to bring the time-to-completion factor under the control of the program.

The most common fixed-pace medium, of course, is film. However, in view of the many similarities between film and a series of slides coordinated with an audiotape recording, the far greater cost of film production did not seem to be justified. The base program, therefore, was planned for presentation by means of an integrated slide-tape recording.

The supplementary elements, however, since they would be prepared independently by a number of programers would not lend themselves easily to the same medium. The logistics of preparing large numbers of slides and tape recordings at the direction of many programers seemed prohibitive. Besides, the task involved in developing such materials was likely to be quite remote from the experience of most of the subjects.

Videotape was selected as the medium for preparing the program supplements because, by providing a stand-up teaching situation, the tasks required of the subjects would be well within the range of experience which might be expected of them, particularly if they were drawn from a population of teachers.

Selection of a Grade Level

The sixth grade was chosen as the locus for the experiment for several reasons. First, since the cost of videotape playback limited the number of times each program could be shown, it was necessary to plan for group viewing of the experimental revisions. Further, while it would have been desirable to randomly assign individual students to viewing groups, it was not likely that schools would cooperate in such a randomization procedure. More likely, playbacks to intact classroom groups would be preferred by schools.

The departmental organization at the junior high and high school levels suggested that classroom groups drawn from that population would be grouped by subject. Multiple-tracking practices in many schools would make additionally difficult the selection of a sample in which all groups had comparable characteristics. Since the project was constrained from pursuing a large sample of groups, it was necessary to achieve as much comparability among the sample groups as possible. It was felt that this would be better achieved at the elementary level where students were not grouped by subject and where the attempt was often made to avoid ability groupings in making up classroom rolls. (These conditions were subsequently checked in the selection of the sample classrooms.)

For ease of administration, too, the elementary level seemed better suited to the purpose. With students assigned most of the day to a single teacher, convenient scheduling of the playbacks promised to be more easily achieved than under conditions where the students were moving frequently between classes.

Sixth grade was selected to achieve the most mature elementary students, since the learning activities promised to involve both reading and writing and it was desirable to reduce the amount of error variance which might be obtained from deficiencies in these skills.

Development of the Experimental Base Program

Selection of the Topic -- Since the experiment involved two procedures for revising a preliminary version of an instructional program, it was necessary at an early stage to develop a base program which subsequently would be experimentally revised.

A number of considerations influenced the selection of a topic for this base program. First, the experiment would benefit from more favorable response from school officials if the total program, which was to be administered in schools, could be administered inside an hour. Therefore, an after-only design seemed desirable as a means to eliminate pre-testing of students within the program. To reduce error variance obtaining from differential foreknowledge of the topic, therefore, it seemed advisable to select a learning task which could be reasonably assumed to be unfamiliar to the vast majority of students.

With no need to pre-test, the administration of the program might be accomplished within the following broad limits:

Administration of base program	15-20 min.
Administration of videotape supplement . . .	15-20 min.
Administration of post-test	10-15 min.
Introduction, instructions, etc.	5-10 min.
Total	45-65 min.

If the base program were to be administered in 15-20 minutes as the above schedule suggested it must, the scope of its objectives had to be severely limited. To test the learning of students in the short remaining post-test interval, the learning tasks needed to be fairly discrete and amenable to efficient, objective measurement. Finally, if we wished later to generalize our results across some larger population of learning tasks, the experiment would benefit by selecting a topic possessing some generalizable traits.

With these considerations in mind, the topic of *English Money* was particularly appealing for the following reasons:

1. Its learning tasks were largely symbolic and involved both verbal and quantitative symbols.

2. We could reasonably assume not only that the topic itself was unfamiliar to most American students, but also that its underlying verbal and quantitative systems would be equally unfamiliar. Not only would the student have to learn new names and associate them with unfamiliar coins, but he would also have to learn to calculate values in a

non-decimal system, based upon a mixture of 12-base and 20-base systems.

3. A large number of unambiguous teaching and testing items could be generated from a relatively few principles.

Once the general topic had been selected, it was necessary to limit the scope of the program dealing with it so that it might reasonably be learned within the short time span allowed. An analysis of the learning tasks was conducted and a set of general objectives established with this limitation in mind.

Analysis of the Learning Tasks -- The student encountering English money for the first time is generally exposed to the coins and bills in circulation, to price tags and labels, and to merchants who quote prices and label pieces of money in colloquial speech. The tasks the student is called upon to perform in the presence of these aural and visual symbols involve computing equivalences and differences within the English money system and expressing the result in terms of several symbol systems interchangeably. A typical purchase might involve a merchant quoting a price in colloquial speech, the selection of an appropriate set of coins by the student to cover the purchase, the computation of the correct value of change to be received, and the assessment of equivalence between the value of the expected change and the value of the change received.

In designing the base program, it was decided to expose the student to the expected range of stimulus classes asking him to make the typical responses he would be expected to make under "realistic" circumstances. Thus, in the presence of a variety of stimulus situations the student would be asked to express equivalences and differences in value in any of the conventional symbol systems.

For administrative purposes, since they would be difficult to record in the post-test situation, spoken responses were excluded from the program.

The Objectives of the Program -- The following general objectives were established for the program (see American Institutes for Research, 2, and Orlando, 22).

1. To be able to recognize and label the seven English coins in common circulation.

2. To be able to calculate equivalences and differences of value between coins, written and spoken verbal statements and numerical expressions interchangeably according to the common practice.

3. To be able to express values in terms of coin selections, written verbal statements and numerical expressions according to the common practice.

Specifically excluded objectives were described as follows:

1. To be able to compute American monetary equivalents of British units.
2. To be able to estimate costs of common items in British commerce.
3. To be able to demonstrate oral mastery of the content.
4. To be able to calculate differences involving values in excess of £1.

Stimulus and Response Classes -- Since it would be difficult to provide all students with a set of actual British coins, it was decided that the program and subsequent test would present photographic representations of the coins and their respective labels as follows: half-penny, penny, three-penny bit, six-penny bit, shilling, florin, and half-crown. The program would present a range of values from ½d. to several pounds expressed in terms of written words, spoken words and numerical expressions. In addition, values from ½d. to 19/11½ would be presented in the form of photographic representations of coins.

The student would be asked to respond to these stimuli by expressing equivalences and differences of value in the form of written verbal statements, numerical expressions, and selections from among alternative sets of coins represented photographically.

Table 1 shows the stimulus and response elements as they were operationally defined in the program.

The Base Program as a "Preliminary Draft" -- An initial draft of the program was written and recorded on audiotape to determine the elapsed time from start to finish. The first version ran over 40 minutes and was subsequently cut to 25 minutes by deleting and collapsing frames. As it was the intent of the base program to represent a preliminary draft version, no prior testing was conducted with students and much of the instruction which was felt to be ineffective was left to stand. The final version of the program ran 25 minutes and consisted of 65 2- by 2-inch slides coordinated with a tape (See Appendix B , p. 71).

Development of Post-Test

Since the post-test was to assess in part the student's ability to recognize oral expressions and statements, it was designed to include both visual and aural stimuli. Thus, a tape recording was developed to accompany a paper-and-pencil test.

The technique of playing a tape recording as part of an achievement test was useful for other purposes as well. By this means it was possible to standardize not only the aural elements of the test

TABLE 1

Operational Definitions of Stimulus-Response Elements Managed in the Base Program

Program Element	As Stimulus	As Response
Coin Configuration	A photograph of coins is used as a stimulus pattern.	S selects from several alternative photos of coins to express equivalence or difference.
Coin Labels, spoken	The names of coins are spoken colloquially to S (eg. "Thrup'ny Bit").	S speaks the names of coins colloquially. However, spoken responses are excluded from this program.
Coin Labels, written	The names of coins are written out correctly spelled (ie. not as colloquially pronounced, eg. "Three-penny Bit").	S writes the names of coins as they would be correctly spelled (ie. not as colloquially pronounced, though spelling is NOT an objective of the program).
Value Statements, spoken	A value is spoken to S colloquially (eg. "Tuppence Hayp'ny").	S expresses a value colloquially. However, spoken responses are excluded from this program.
Value Statements, written	A value is written in words (eg. "Two Pence Half-Penny").	S expresses a value in words by writing them as a response (eg. "Two Shillings").
Value Expressions	A value is expressed in terms of numerical symbols according to the conventional practice (eg. "12s.1d=12/1").	S expresses a value in numerical symbols according to the conventional practice (eg. "6/5=6s.5d.").

but also the timing of the test. Further, it was decided to use the recording to read all the test items as a means of reducing the error variance likely to accrue from expected reading deficiencies at the sixth-grade level.

Thirty test items were developed and recorded using a representative sample of the item-generating statements shown in Table 2. A trial of the base program was conducted in two randomly selected sixth-grade classes and the 30 items were administered as a test following the program. Student responses to these items were analyzed by computing indexes of difficulty and discrimination for each item. The most difficult and least discriminating items were thus eliminated in developing the final ten-item test.

The initial 30-item sample is shown in Appendix A, p. 57, together with the indexes of difficulty and discrimination obtained for each item. The final version of the post-test appears in Appendix B, p. 82.

Standardization of Treatments

The task assigned to all the programer-subjects (S) in the experiment, was within a single day, to examine a preliminary version of a 25-minute lesson on English Money consisting of a set of 2- by 2-inch slides coordinated with a tape recording, then to review a set of instructions which included statements of the program's objectives, and finally to plan and record on videotape a 15-minute supplementary lesson calculated to help sixth-graders score well on a post-test following the program.

To assure that comparable conditions prevailed for all S's, standard procedures were followed in recruiting subjects for the experiment, introducing them to their task, producing their videotaped supplementary lessons, and administering these programs in classrooms.

Recruitment -- The following advertisement was placed in a local newspaper:

> Teachers needed for TV teaching project on UCSC campus. One day only. Call . . .

During the subsequent telephone interviews, the experimenter (E) determined if the callers possessed the minimum qualifications for participation. With allowances for interacting with each caller informally, the project was described in the following standardized terms:

> We are engaged in a study concerned with television teaching, particularly with the kinds of information teachers need to help them teach effectively by television.

TABLE 2

Behavioral Objectives Expressed as Test-Generating Statements

S is given any of the following:	And is asked to subtract any of the following:	And to express his answer by any of the following:
- The names of some coins (written out).	- The value of some coins named in writing.	- Naming a correct set of coins in writing.
- The names of some coins (spoken aloud).	- A given amount of money spoken aloud.	- Stating in words (written) a correct amount of money.
- A set of coins (in photographic form).	- A given amount of money written out in words.	- Expressing a correct amount of money by writing a conventional numerical expression.
- An amount of money (written as a conventional numerical expression).	- An amount of money written as a conventional numerical expression.	- Selecting a correct set of coins from several sets shown in photographs.
- An amount of money (written in words).	- The value of a given set of coins (in photographic form).	(Oral responses excluded)
- An amount of money (spoken aloud).	- The value of some coins named aloud.	

> If you are selected as one of the participating teachers, this is what you would do: we would set up a time when it was convenient for you to spend a day with us at our studios on campus. At that time you would be assigned a topic and asked to prepare a 15-minute lesson on that topic which would be presented in the studio and recorded on videotape, all on that same day. Later we would play back your videotape in one or two schools in this area to determine how well students learned from it.
>
> When you arrive here, we will assign you a topic and then show you around the studio and demonstrate the tools and techniques you can use to teach it. You will have most of the day to plan your lesson and try it out before recording it. My staff and I will be there to help you at all times.
>
> Most of our participants have had no prior experience with television, so we are prepared to teach you the basic fundamentals. Also, since this is a recording, you can prepare one small part at a time, review it, and then do it over as many times as you wish until you are satisfied.
>
> This is basically a research project, so no publication of the videotape is planned. If you are interested in television teaching, this is a good way to start. If you decide you would like to join with us, the project is prepared to compensate you $20.00 for your participation.

Introduction to the Programing Task -- It was anticipated that most subjects would have reason for feeling insecure when faced with the task that was proposed. Few had any experience with television in the first place, which accounted for their initial interest in the project. Experience had indicated, however, that the first exposure to the hardware and staff associated with television would arouse feelings of insecurity. Further, for most subjects, the topic of English Money was certain to arouse additional feelings of tension -- it was a topic calculated to be unfamiliar to American audiences. The introduction, therefore, while concerned with standardizing the experimental conditions for all subjects, was equally concerned with relaxing each subject and providing a basis for feeling confident. The following discourse was used as a guide, therefore, and every opportunity to gain the subject's participation in a dialogue was exploited:

> We are preparing a lesson for sixth-graders on the topic of English Money. The entire lesson will consist of several elements. First, we have prepared a 25-minute sequence of slides with an accompanying tape which introduces the subject. It is written after the fashion of self-instructional programs, so the students are given a chance to practice what they learn as they go along. You will have a chance

to see this sequence, and go through it just as the students will. That way you will see exactly what the students will see and, at the same time, learn about English Money yourself.

The second part of the lesson will be a 15-minute videotape supplement to the slide sequence, which you will prepare today. The object of this videotape is to provide whatever additional instruction or practice you think the students will need to help them pass a test on the subject.

Finally, the students will be given a brief test on the subject to see how much they have been able to learn from the entire lesson -- the slide sequence, and your videotape supplement.

We have prepared this booklet of materials [see Appendix B, p. 65] to help you in your task. It contains, here, a general description of the objectives which the lesson will try to achieve. Here is a table which defines the stimulus-response terms operationally. Next, here is what might be termed a "test-item generator." This table shows what kinds of test items will appear in the final test of the subject. Following, here, is a copy of the script which was used for the slide-tape sequence so you can review it easily. [Treatment II adds: Finally, here is a copy of the specific items used in the post-test together with data showing how well sixth-graders perform on each item after seeing just the slide-tape sequence.]

Your job is simply this: study the slide-tape sequence and the objectives [Treatment II adds: . . . and the test results]. On the basis of your own experience and knowledge about how people learn, try to plan a lesson which will help the students pass the final post-test. [Treatment II adds: The only thing we must ask, of course, is that you do not use any of these specific test items in your presentation.] You will have six hours in which to complete your tape. You should plan to spend no more than three hours in planning before you start to practice in the studio.

At this point, S was given a tour of the studio which demonstrated the studio facilities and the operations of which they were capable. In particular, the capabilities of the zoom lenses were demonstrated as well as applications of the overhead camera. It was pointed out that their lesson could be expository or it could involve student practice as they wished.

Following the introduction to the studio, S was seated in the planning area and the slide-tape sequence was started. From then, S was on his own to plan his lesson. The technical staff remained in the area, out of sight in the control room, to provide technical assistance in the studio if S wished it.

The Physical Arrangements -- A standard studio setup was provided to all S's and all were limited to using only the resources provided in the studio. The standard studio setup, shown in Figure 4, provided the following complement of facilities:

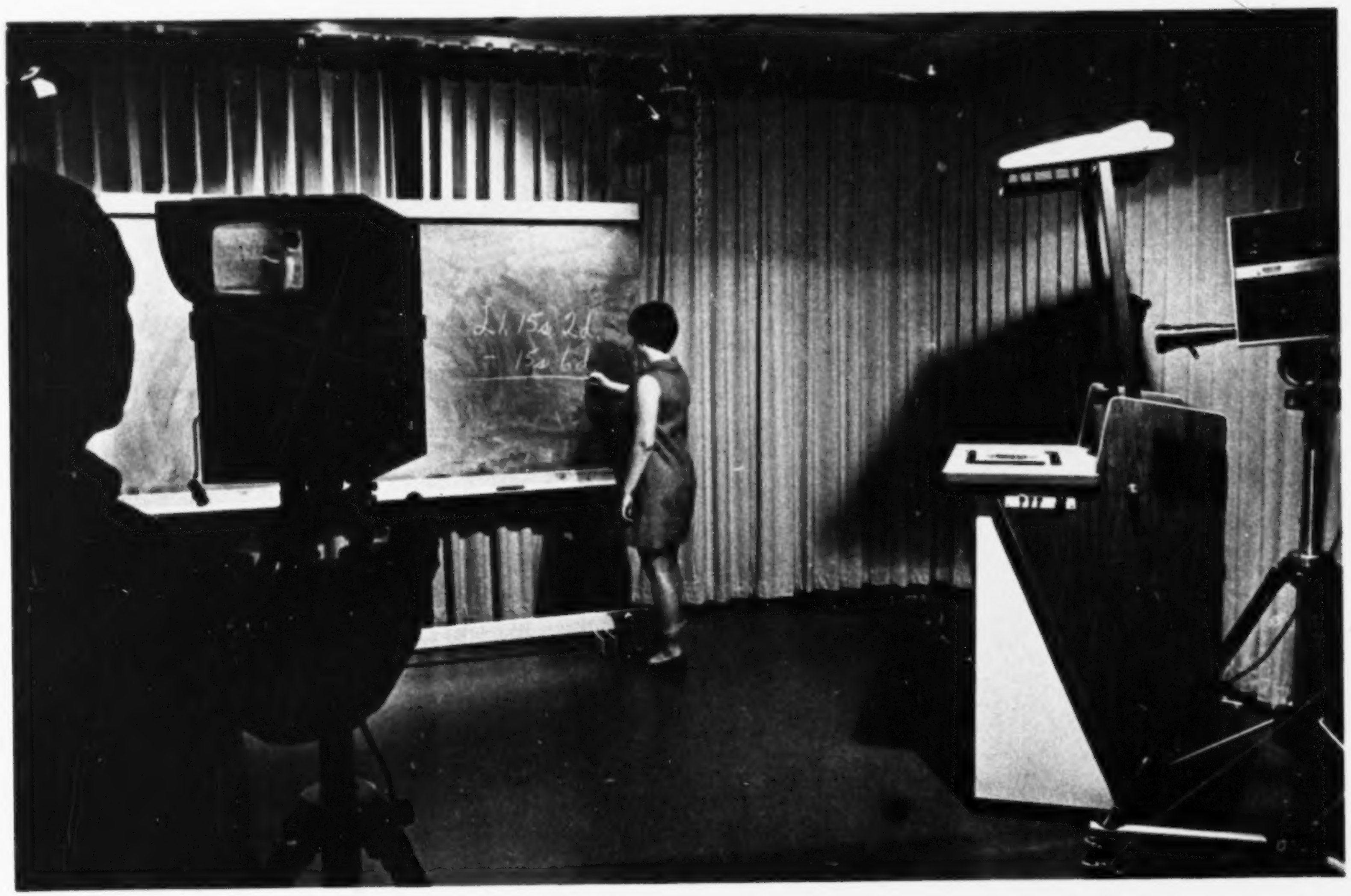
1. A fixed camera with 15-150 mm. zoom lens
2. A movable camera with operator and 15-150 mm. zoom lens
3. A subject-operated overhead camera system with 15-150 mm. zoom lens
4. Two studio monitors
5. A chalkboard
6. A world map
7. A globe
8. Miscellaneous furniture as needed.

A separate area of the studio was set up with a desk and rear-projection slide viewer for displaying the base program and planning the supplementary lesson and graphic materials, shown in Figure 5. The total complement of items provided to S in this area consisted of the following:

1. An instruction booklet for either Treatment I or Treatment II (see Appendix B, page 65.)
2. A set of English coins (one half-crown, two florins, three shillings, five six-penny bits, four three-penny bits, seven pennies, and four half-pennies).
3. Blank paper and felt-tipped pens.

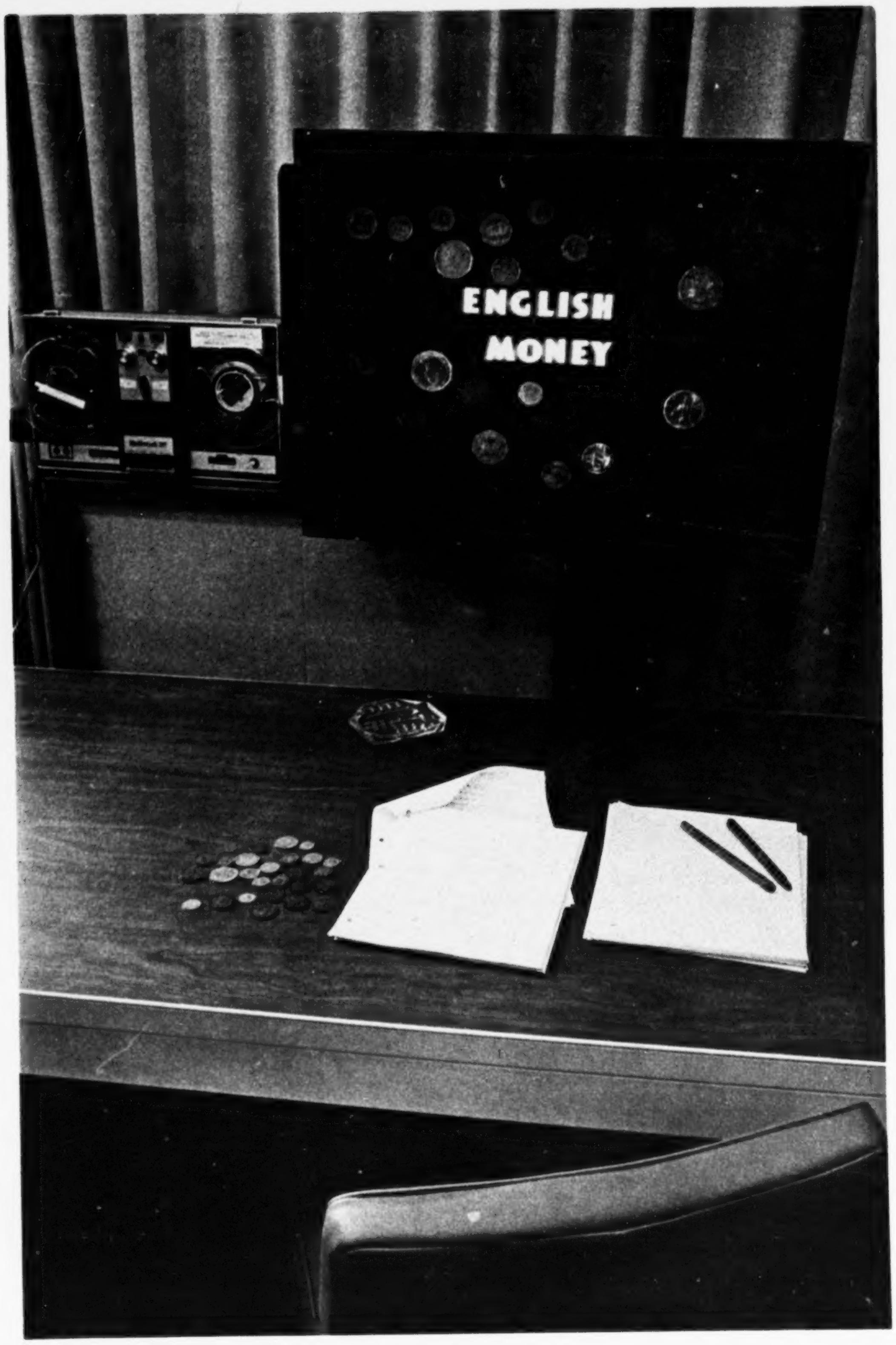
General Constraints -- Several constraints were imposed upon all these activities to assure comparability of treatments. First, no S was informed of the topic to be taught at any time prior to his arrival at the studio on the day of production. In this way, E could assure that S's were not able to prepare for the production in any differential ways. Particularly, the possibility that S's with later appointments would gain an advantage by being able to prepare longer was avoided.

FIGURE 4


£1. 15s 2d
− 15s 6d

Standard studio setup for production of supplementary videotaped lessons.

FIGURE 5


ENGLISH
MONEY

Standard setup for lesson planning and preparation.

Further, technical studio personnel were constrained from offering suggestions or interacting with the S's in any other than technical matters. Consequently, there was no "directing" of the subjects in the usual production sense, nor was there interaction with S which might bias strategic decisions in planning the lesson. The S's were left totally responsible to plan the content of their lessons and the technical staff instructed to confine their assistance to technical matters entirely.

Finally, the exact nature of the variables under consideration was withheld from the subjects as well as from the technical staff throughout the experiment. In this way, whatever effects might result from the interaction of the S's with the television staff might be randomly distributed across treatments.

From the time of arrival until final completion of the videotaped lesson all S's conducted their work entirely within the immediate studio area without leaving the premises.

Introduction to the Classroom Trials -- A standard procedure was followed in the trial classes to assure that all programs were administered under comparable conditions.

Prior to the trial, it should be noted that no teachers were advised of the subject of the program so that all classes first learned of the topic as presentation of the lesson was started. School administrators had to be informed of the topic in advance, however, in order to gain their approval for presenting the program in the schools. In all cases, the importance of avoiding prior preparation on the subject in the trial classes was emphasized. The procedure followed by them was to advise the teachers only that a "worthwhile subject" would be taught. In every case the possibility of prior preparation was examined in the introductory procedures to assure comparability in this area. In no case was prior preparation in evidence.

The following brief introduction was given to each class immediately prior to starting the lesson:

> Hello, boys and girls. My name is _______________ and I work for the University of California. I have a special lesson for you today which is such a surprise, even your teacher doesn't know what it is about.
>
> We are making a filmstrip . . . is there anyone here who doesn't know what a filmstrip is? Right, it's a series of pictures that teach a lesson, isn't it? . . . Well, we are making a filmstrip on the subject of English Money and you are going to help us find out how good a job of teaching our filmstrip is doing.

Some of you may already know something about English money. Who remembers why English money was in the news recently? (Field responses which may cover devaluation of the pound, and which may name some of the denominations of English money. Make sure prior knowledge is limited.) Good. Well, today we are going to learn about the denominations of English money, the names of their coins and how to make change in the English system.

Now this is a different kind of a lesson. This filmstrip has a tape that goes along with it, so you not only watch the pictures, but you listen to the teacher as well. The teacher will ask you some questions as you go along, and you should try to write your answers to these questions on this answer sheet. It will give you practice in what you are learning. (Pass out worksheets.) Let's look at this worksheet. You can see there are a lot of empty boxes on the right side of the sheet, numbered from 1 to 35. And there is one marked "A" just before the first box. That one is for practice. During the lesson, as the teacher asks you questions you are to try to think of an answer to each question and write it down in the answer space. Remember, this is just a worksheet, not a test. We're going to throw these away afterwards. Notice the scratch area on the left. That is for the arithmetic you might have to do in figuring up your answers.

Right after you have seen the filmstrip, we have prepared a review lesson on television. So you will have a chance to review what you have learned. [See note below.]

Finally, after the review lesson, we will have a short test.

But remember, this is not a test to see how well you do. It's to see how well I have done at teaching you. If I have done a good job of teaching, then I'll know this filmstrip is a good one because you learn from it. But if you don't learn too well, then I know I haven't done a very good job and I will take the filmstrip back to the university and try to make it better.

In all the things we will do in the next hour, you will not put your names on anything. This will not be part of your grade. So you can just relax, and enjoy yourselves, and see how much you can learn about English money.

To review: First, the filmstrip lesson with the worksheet, then the television review, then a test. Any questions? (Answer procedural questions.) Then let's start the tape. If you need a pencil, raise your hand. Otherwise, once we start, we cannot stop. Do your best, and try to keep up with the lesson at all times. Here we go. (Start base program.)

(After base program.)

All right, now turn your worksheet over so you have a blank sheet of scratch paper for the television review. Try to fix in your mind what you need to know to answer the questions on the final test. (Start videotape program.)

(After videotape program: Collect all worksheets, pass out test booklets.)

Remove the blank cover sheet on the test booklet. Use this to cover your answers (demonstrate) and for scratch paper if you need it. Let's look at the test. These empty answer spaces on the right are for your answers. Please put them there so I won't miss any when I correct the papers.

I have prepared a tape to go along with this test. Each question is read to you, and then a certain amount of time is given for you to answer the question before going on to the next question. Do the best you can, and try to keep up with the tape. You can go back if you want to, but when the tape is over, the test will be over. Questions? (Field procedural questions only.) (Start tape.)

(After test, collect papers; thank all.)

Note: All references to the television review were omitted for the base program trials.

Administration of Program Trials

As supplements to the base program, 20 videotaped lessons were prepared by 20 programers representing two independent sample groups following different procedures. Representative scenes from these programs are shown in Figures 6-25. To assess the relative effectiveness of each procedural group, measures of effectiveness were needed for the videotaped lessons produced by the programers assigned to each group.

The strategy employed called for a trial of each videotaped lesson to be conducted in one randomly assigned, intact class. The post-test was to be administered following the lesson and the mean of the scores obtained by that class, expressed as a percentage of possible score, was to serve as a measure of the lesson's effectiveness.

The 20 trials of the videotaped lessons were conducted in 20 randomly assigned sixth-grade classes over a period of three months. In addition, during this same period seven trials of the base program, without supplements, were conducted in seven randomly assigned sixth-grade classes.

FIGURE 6

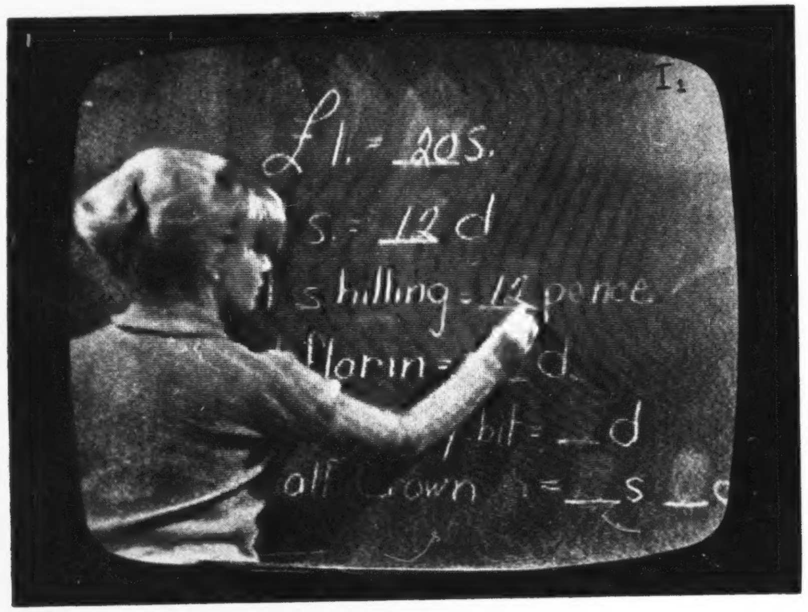

I₁
£1. = 20s.
s = 12 d
1 shilling = 12 pence
florin = __d
bit = __d
alf crown = __s __d

FIGURE 7

GLASGOW
UNITED
BRITISH
NOR
ISLES
NEWCASTLE
LIVERPOOL
KINGDOM
SEA
MANCHESTER
English Channel

FIGURE 8

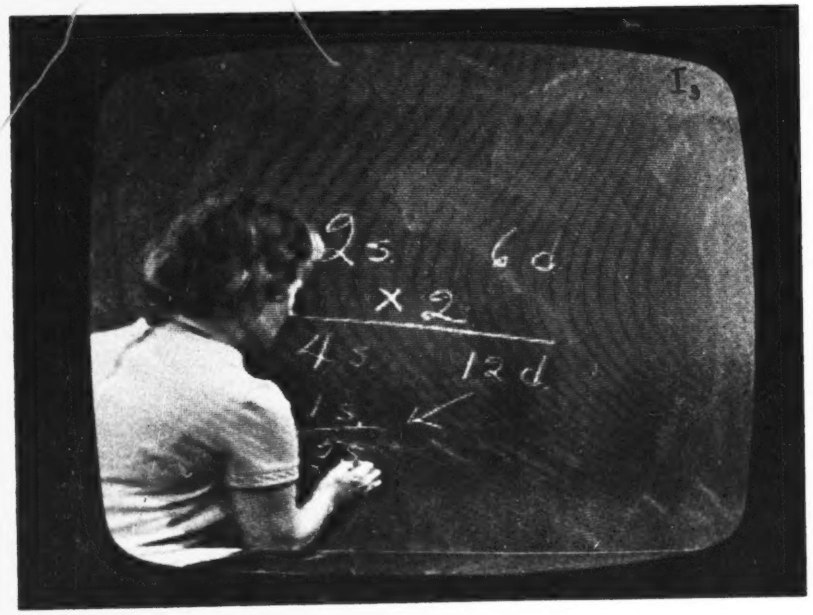

I₃
2s 6d
× 2
4s 12d
1s

FIGURE 9

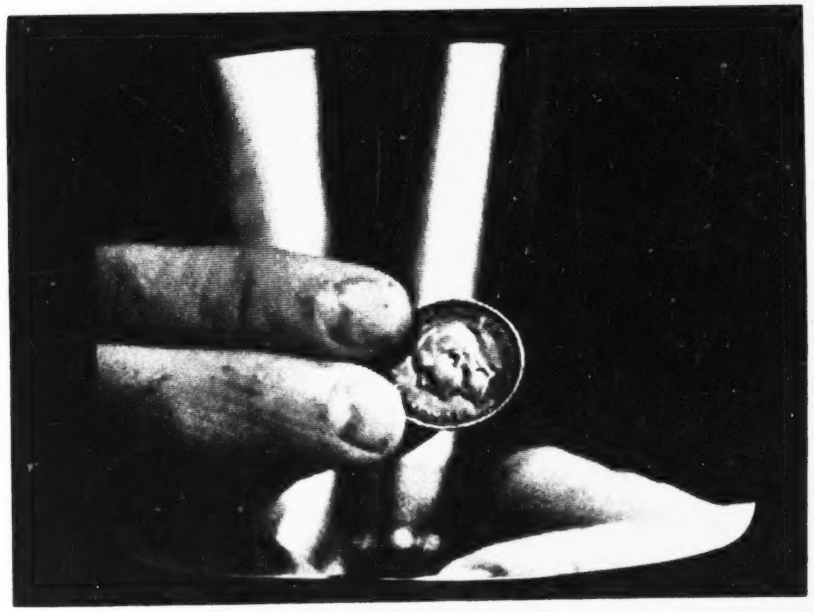


FIGURE 10

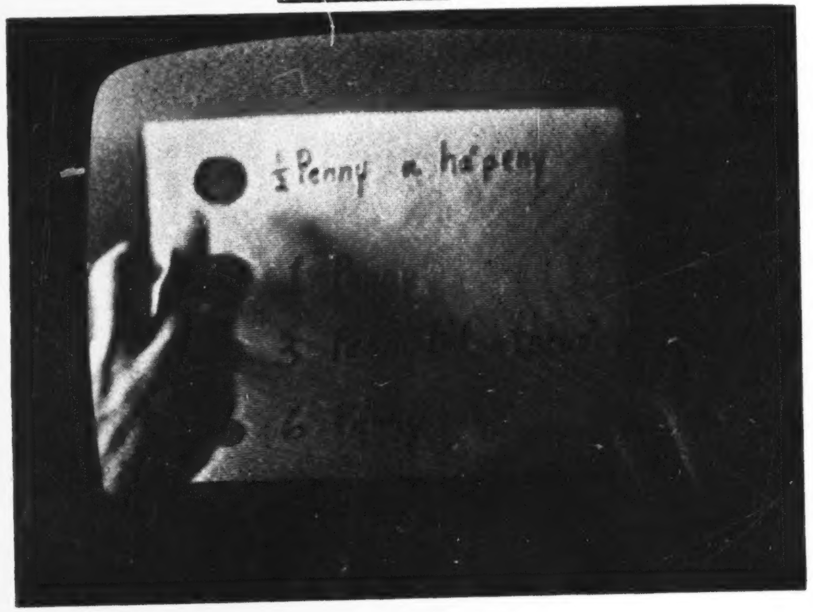

½ Penny a hapeny

FIGURE 11

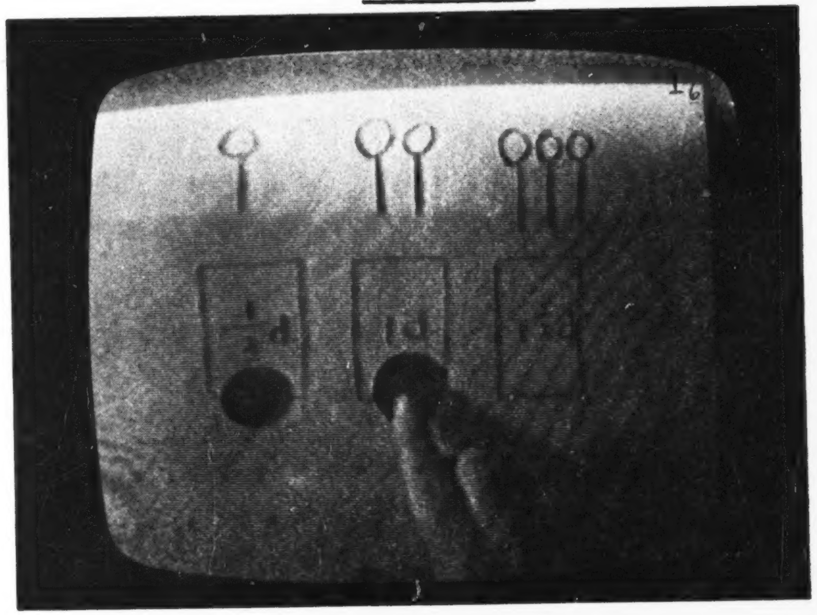


FIGURE 12

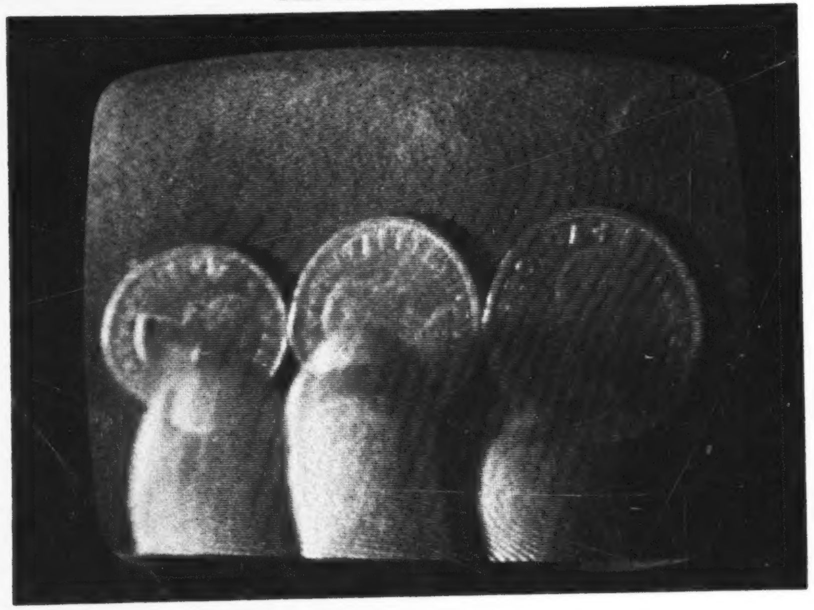


FIGURE 13

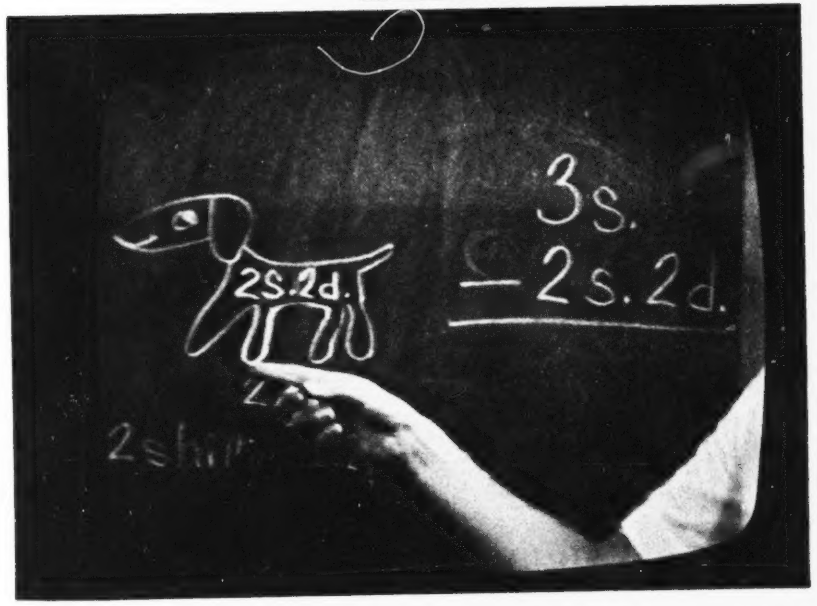

3s.
−2s.2d.
2s.2d.

FIGURE 14

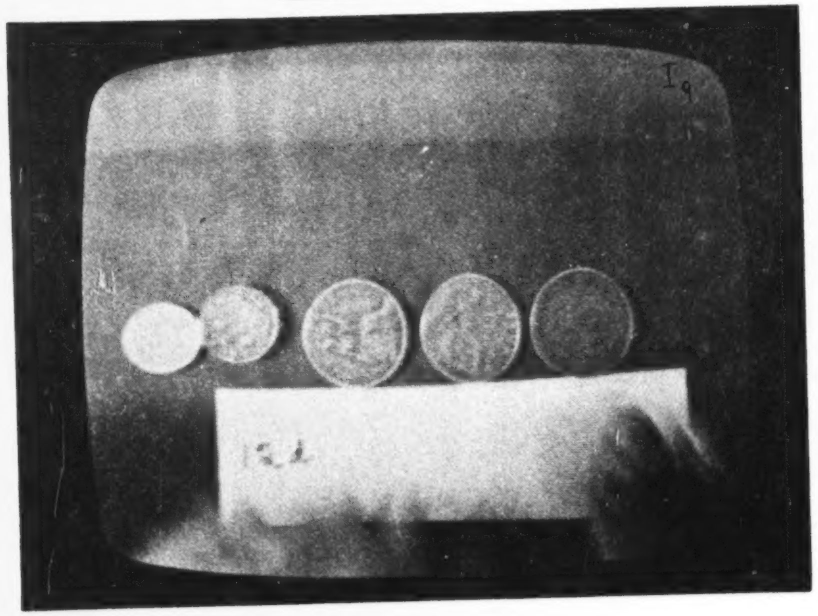


FIGURE 15

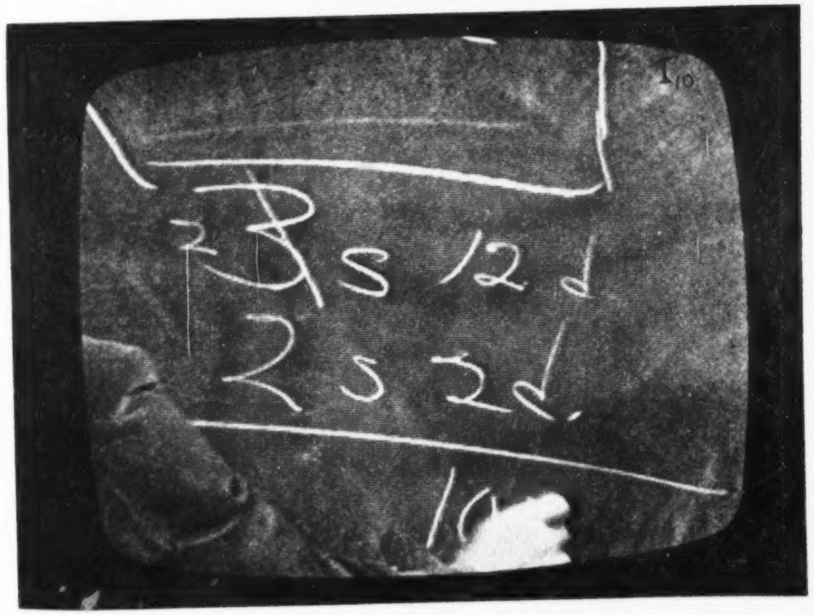

I10
23s 12d
2s 2d.
10

FIGURE 16

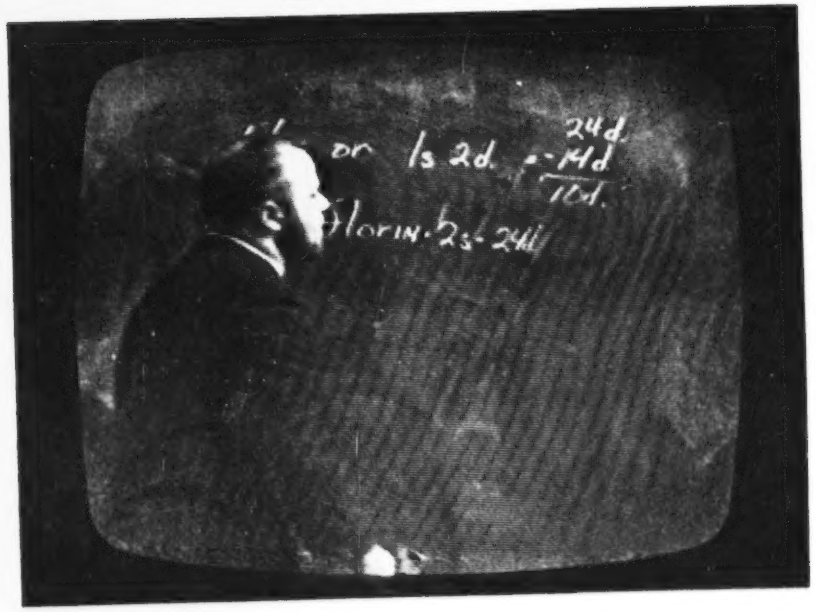

or 1s 2d.
24d
-14d
10d.
Florin. 2s. 24d

FIGURE 17

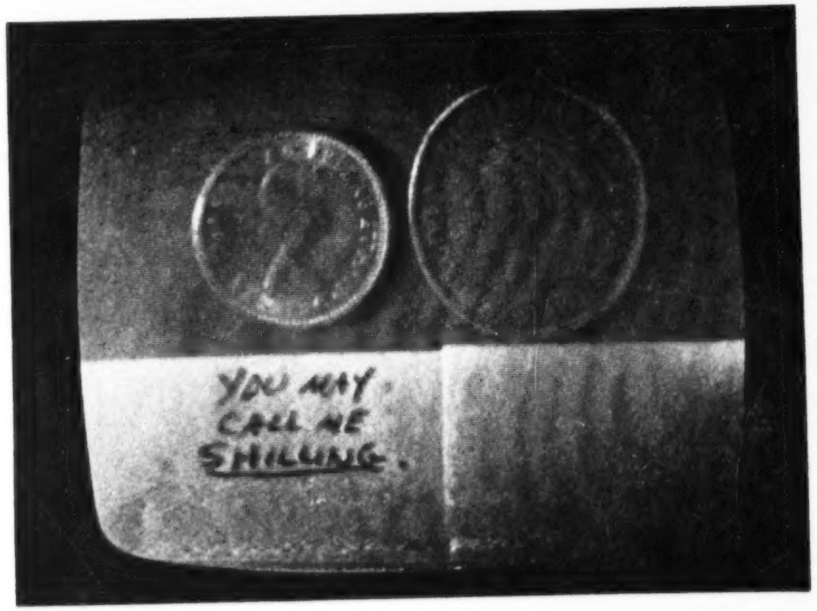

YOU MAY
CALL ME
SHILLING.

FIGURE 18

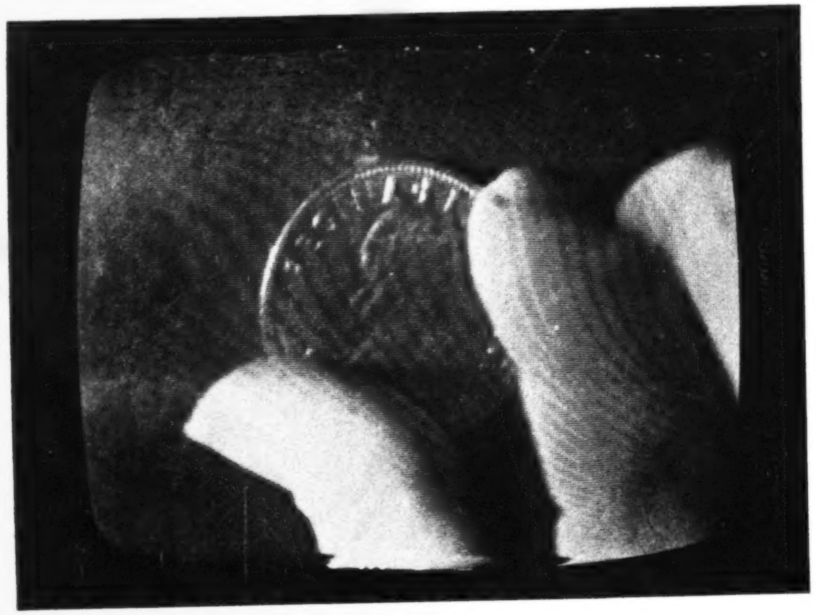


FIGURE 19

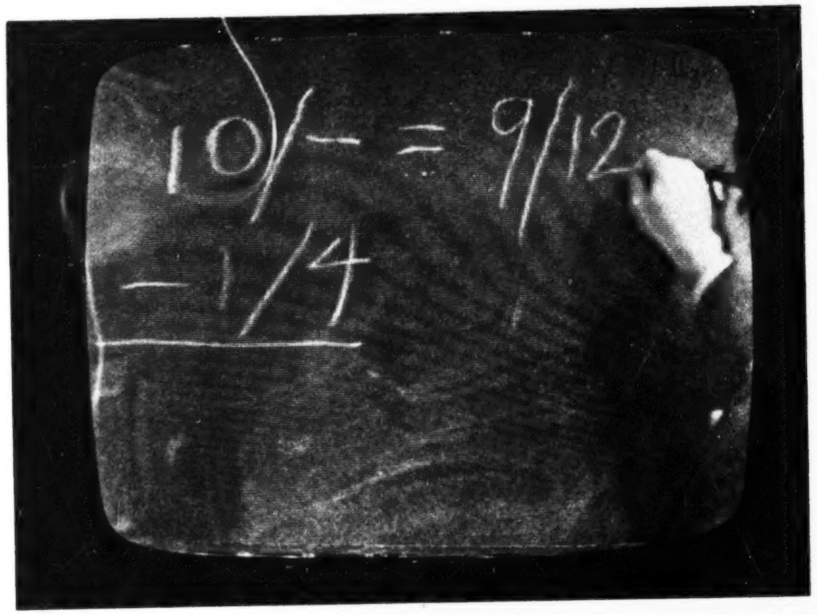


FIGURE 20

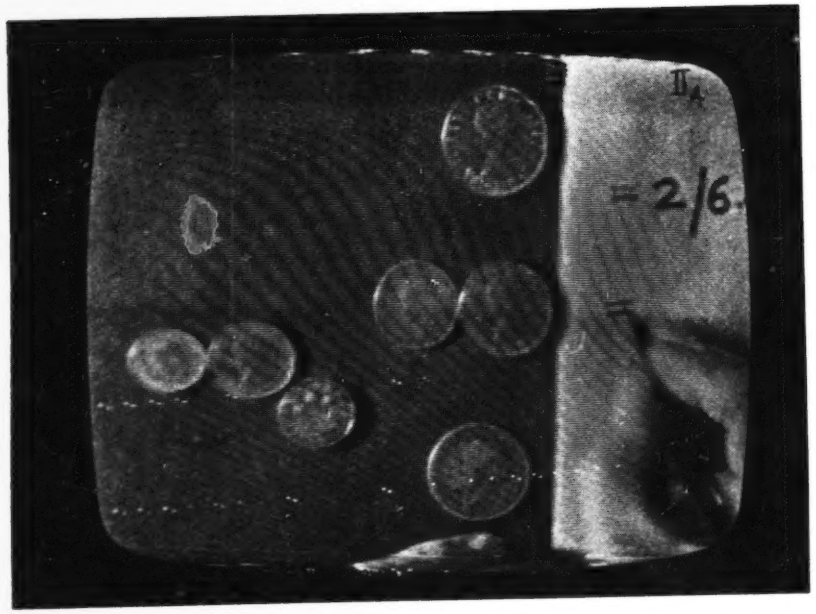
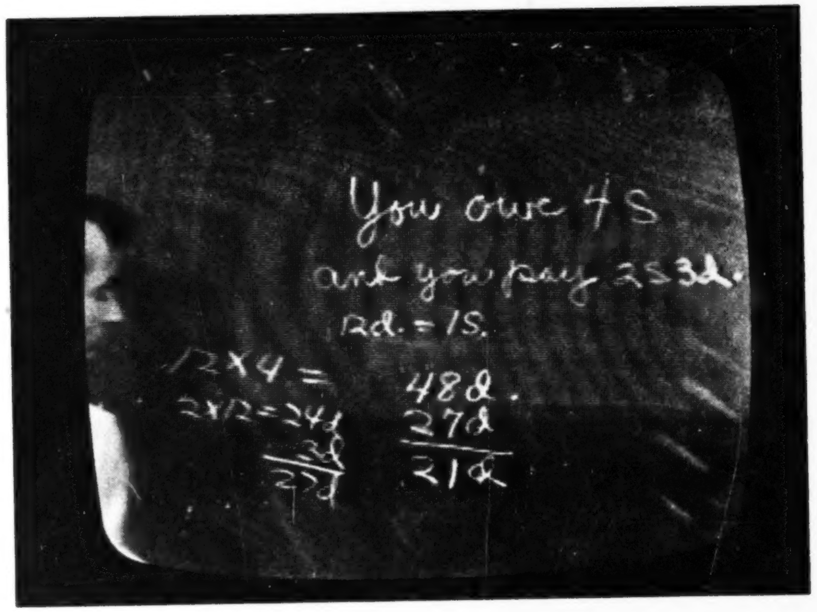


FIGURE 21



You owe 4S
and you pay 2S 3d.
12d. = 1S.
12×4 = 48d.
2×12=24d
3d
27d
27d
21d

FIGURE 22

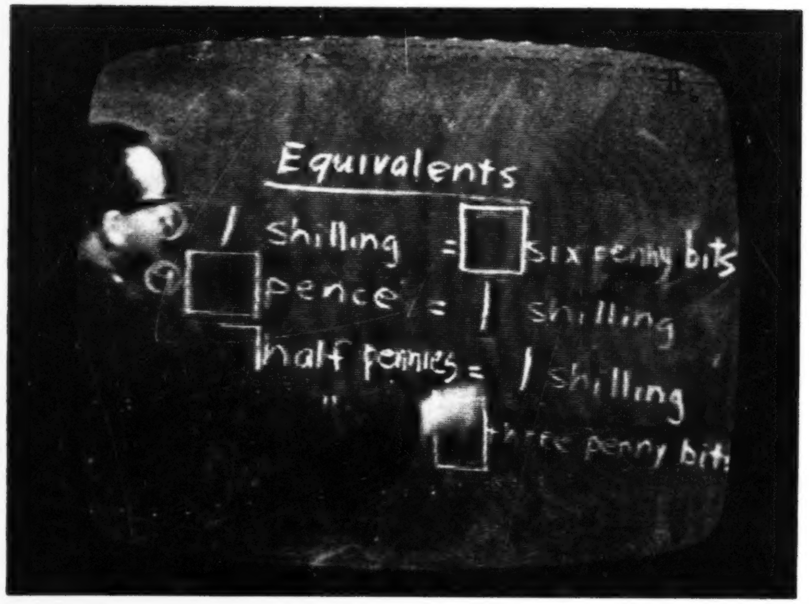

Equivalents
1 shilling = ☐ six penny bits
☐ pence = 1 shilling
half pennies = 1 shilling
" ☐ three penny bits

FIGURE 23

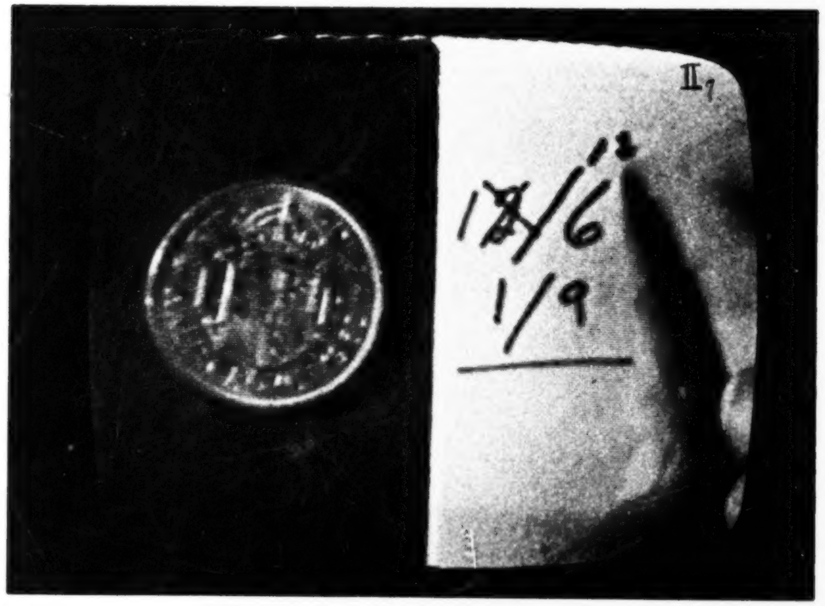


FIGURE 24

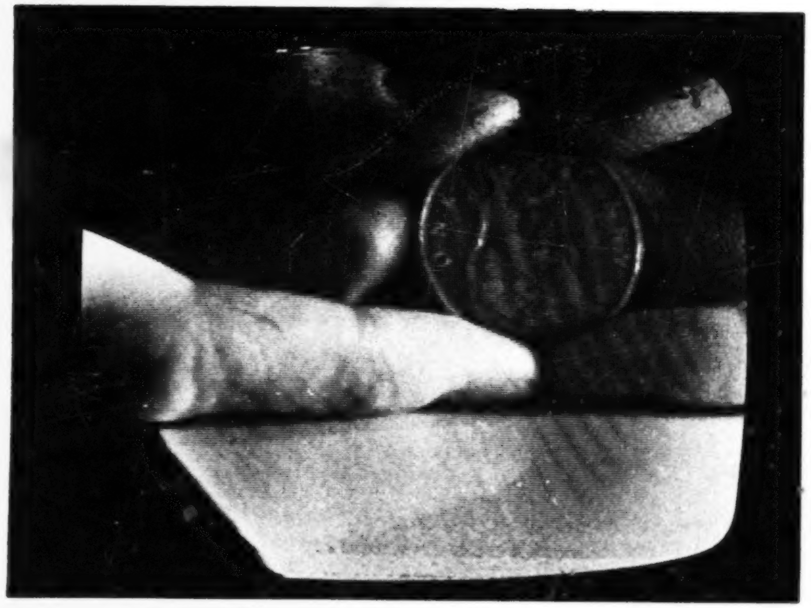


FIGURE 25

ANALYSIS AND FINDINGS

To determine the relative effectiveness of the original version of the program (orig) and the revisions produced under two treatment conditions (revI and revII), the results of an identical test administered to all trial classes were analyzed. The means (M) of the class means for all treatments were analyzed to test the significance of their occurrence in a predicted order of magnitude under H_0, i.e., to test H_0: $M_{orig} = M_{revI} = M_{revII}$ against the specific alternative H_1: $M_{orig} < M_{revI} < M_{revII}$.

Problems of Analysis

The experimenter was able to state, at the outset, the expected rank order of magnitude of the effects of the programs prepared under the different experimental treatments. It was expected that, of all the programs, the original, unrevised version would occasion the lowest achievement scores, that the revised version prepared by subjective analysis of behavioral objectives (revI) would, as a group, occasion higher scores, and that the revised version prepared by objective analysis of test scores obtained after a trial run of the original (revII) would, as a group, occasion the highest scores. A test of significance was desired especially sensitive to those differences which, while tending to reject the null, at the same time would lend support to the specific alternative H_1, that the magnitudes of the treatment effects would occur in a particular order.

It is known that the customary one-way analysis of variance does not satisfy this demand: the F-ratio, as well as the H-ratio and other k-sample statistics are independent of the order in which the sample means occur. Furthermore, attempts to assess the joint probabilities yielded by F and some coefficient of rank correlation between an expected and an observed order of means meet with a number of difficulties (see Jonckheere, 8).

Thus we had the problem of determining if the results obtained from a one-way analysis of variance design supported an alternative hypothesis that the magnitude of the effects produced by the different treatments occurred in the predicted order. When only two treatments are under consideration, the hypothesis that the means will occur in a specified order is taken into account by the usual application of the one-tail test of significance. If E is unconcerned with the direction of difference, the two-tail test is used. A procedure which extends the notion of a predicted ordered alternative to the k-sample case is not in common use.

Jonckheere's Procedure

Jonckheere (8) proposes a procedure whereby the statistic S may be used for testing the null hypothesis against the alternative of ordered cumulative distribution functions according to the following:

$$S = 2\sum_{i=1}^{k-1} \sum_{j=1+i}^{k} p_{ij} - \sum_{i=1}^{k-1} \sum_{j=1+i}^{k} m_i m_j \qquad (1)$$

where p_{ij} is obtained by arranging the samples in the order implied by H_1 and then, for each sample in turn, determining for each value the number of items which are larger in all succeeding samples.

The value of S yielded by this procedure he shows to be precisely the same as that obtained in the calculation of Kendall's S between two rankings, when one of them contains ties. In view of the identity of the statistic employed in this test and Kendall's S, Jonckheere shows that for large samples

$$z = \frac{S-1}{\sqrt{\left[\frac{1}{18}\left\{n^2(2n+3) - \sum_{r=1}^{k} m_r^2(2m_r+3)\right\}\right]}} \qquad (2)$$

is approximately normally distributed with zero mean and unit variance. Therefore, the probability associated with the occurrence under H_0 of any observed value of S may be determined by computing the associated value of z and then determining the significance of that z by reference to a standard normal curve table showing the probabilities associated with various values of z.

It was proposed then to test H_0: $M_{orig} = M_{revI} = M_{revII}$ against the specific alternative H_1: $M_{orig} < M_{revI} < M_{revII}$ by calculating S as shown by Jonckheere and then the associated value of z, testing the significance of the observed value of z.

Analysis of Data: H_0 vs. H_1

The data we wish to analyze are shown in Table 3. It will be noted that the samples are arranged in the order of predicted increasing magnitude and that the means are, in fact, in the predicted order. We wish to test the hypothesis that the three samples have come from the same population against the alternative that the populations are such that the values of the three samples are in the expected order of increasing value.

For the computation of S we have

$m_1=7$, $m_2=11$, $m_3=9$, $k=3$,

$p_{12}=54$, $p_{13}=55$, $p_{23}=78$

Hence, using (1) we have

$$S = 2(54 + 55 + 78) - 239$$

$$= 135$$

Now computing the associated z using (2) we have

$$z = \frac{135 - 1}{\sqrt{\left[\frac{1}{18}\left\{(27)^2(57)\right\} - \left\{(7)^2(17)+(11)^2(25)+(9)^2(21)\right\}\right]}}$$

$$= \frac{134}{44.72}$$

$$= 2.996$$

By reference to a normal distribution table we find that $z \geq 2.996$ has a probability of occurrence under H_0 of $p \leq .0013$. Thus, we reject H_0 that the samples were drawn from the same population and accept the specific alternative H_1: $M_{orig} < M_{revI} < M_{revII}$, that the samples were drawn from populations with ordered cumulative distribution functions as expected at a level of significance of $p < .0013$. This analysis is summarized in Table 4.

The cumulative distribution characteristics tested by this procedure are shown graphically in Figure 26. Reference to the figure shows various values of the cumulated sums of class means arrayed along the axis of ordinates. Along the axis of abscissas are arrayed the various ranks by which any given class mean may be identified within its own sample. Reference to Table 3 shows the cumulated sums of class means relative to the within-sample ranks. For example, to plot the point associated with the lowest class mean (rank 1) in sample_{orig}, we plot (1, 17.78). The next point in sample_{orig} (rank 2) adds the class mean 18.65 to arrive at the cumulated sum at rank 2 of 36.43. The second point is thus plotted (2, 36.43). By this process the cumulative distribution characteristics of the class means are plotted for each sample group. It can be seen that these functions occur in the order of magnitudes predicted under H_1.

TABLE 3

Summary of Ranked Class Means From Trials by Treatment

Original Program (orig)			Subjective Revision (revI)			Objective Revision (revII)		
Program	Class Means	Cumulated Sum of Means	Program	Class Means	Cumulated Sum of Means	Program	Class Means	Cumulated Sum of Means
$orig_5$	17.78	17.78	$revI_7$	18.75	18.75	$revII_8$	22.67	22.67
$orig_7$	18.65	36.43	$revI_2$	19.00	37.75	$revII_1$	25.42	48.09
$orig_4$	19.33	55.76	$revI_6$	21.92	59.67	$revII_9$	26.25	74.34
$orig_6$	20.89	76.65	$revI_1$	22.31	81.98	$revII_4$	29.48	103.82
$orig_1$	21.50	98.15	$revI_3$	23.29	105.27	$revII_6$	29.81	133.63
$orig_2$	24.23	122.38	$revI_4$	24.17	129.44	$revII_5$	30.91	164.54
$orig_3$	31.83	154.21	$revI_{10}$	24.35	153.79	$revII_3$	31.00	195.54
			$revI_8$	24.48	178.27	$revII_2$	33.39	228.93
			$revI_{11}$	26.59	204.86	$revII_7$	33.70	262.63
			$revI_5$	31.41	236.27			
			$revI_9$	31.80	268.07			
$M_{orig} = 22.03$			$M_{revI} = 24.37$			$M_{revII} = 29.18$		

*Each of 27 programs was randomly assigned to an intact class for trial. Each datum above consists of the mean of percentage scores obtained by that class on the post-test administered after the program. Accordingly, the data were statistically treated using the class rather than the individual student as the unit of observation, with the class mean serving as the measure of class performance.

TABLE 4

Summary of Analysis of Ordered Differences Among Sample Means

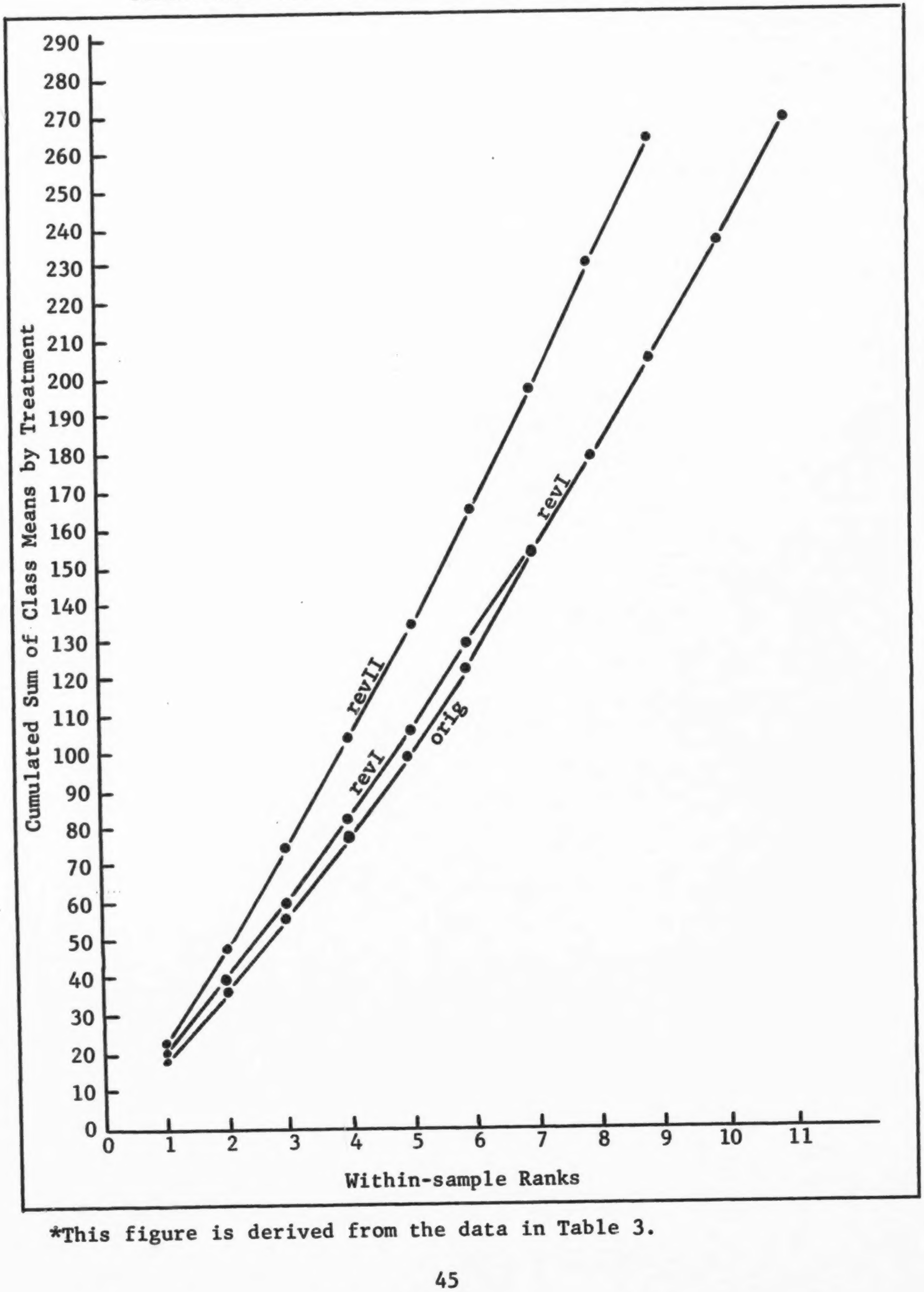
	Original Program	Subjective Revision (I)	Objective Revision (II)
Means of Class Means	22.03 N = 7	24.37 N = 11	29.18 N = 9

H_0: $M_{orig} = M_{revI} = M_{revII}$ was tested against the alternative H_1: $M_{orig} < M_{revI} < M_{revII}$ by means of a procedure described by Jonckheere (8) using the statistic S. In the case of k = 3 and N = 27, the observed value of S = 135 has a probability of occurrence under H_0 such that

H_0 is rejected in favor of H_1 at a level of significance of $p < .0013$.

FIGURE 26

Cumulative Distributions of Class Means by Treatment*



*This figure is derived from the data in Table 3.

To confirm this finding in the more traditional fashion, the Kruskal-Wallis one-way analysis of variance by ranks procedure (Siegel, 25, 184ff.) was applied to test H_0: $M_{orig}=M_{revI}=M_{revII}$ against the more general alternative that $M_{orig} \neq M_{revI} \neq M_{revII}$ yielding H = 15.03 with df = 2. The probability associated with this value of H allows rejection of the null at $p<.001$ and acceptance of the hypothesis that the samples were drawn from different populations.

Further two-sample comparisons were performed using the Mann-Whitney U Test (Siegel, 25, p. 116ff.) to assess the reliability of the differences. With regard to the hypothesis that $M_{orig} < M_{revI+revII}$ it was found that the revised versions, irrespective of which procedure was used, were reliably more effective than the original version at a one-tail level of significance of $p< .025$. Furthermore, with regard to the hypothesis that $M_{revI} < M_{revII}$, the major hypothesis dealt with in the experiment, it was found that the revisions produced following the objective procedure were reliably more effective than the "subjective" revisions at a one-tail level of significance of $p< .025$.

When we test the hypothesis that $M_{orig} < M_{revI}$, however, our results are not nearly so conclusive. Here we find, while the obtained difference occurred in the predicted direction, we are unable to rule out that this difference resulted from chance fluctuations. Thus, if we examine only the data deriving from tests of the original program and the "subjective" revision, we are unable to reject H_0: $M_{orig} = M_{revI}$. It is clear from this analysis, as it is graphically in Figure 26, above, that the bulk of the variance obtaining from tests of all versions of the program can be accounted for by the difference between the effects of revision II and the effects of all other versions.

Accordingly, the hypothesis that $M_{revII} = M_{orig+revI}$ was tested against the alternative that $M_{revII} > M_{orig+revI}$. It was found that the "objective" revisions were reliably more effective than both the original versions and the "subjective" revisions, taken as a group, at a one-tail level of significance of $p< .01$.

To review, with respect to the major experimental hypotheses tested in the experiment, it was found that the effects obtained from trials of the original program and the experimental revisions occurred reliably in the expected order of magnitude. The absolute magnitudes of the differences obtained were such that a more reliable difference was found between the effects of the "objective" revisions and the "subjective" revisions than between the "subjective" revisions and the original version when these differences were analyzed separately. Yet, when taken together, support was found for the hypothesis that both revision procedures would improve upon the original, the "objective" procedure to a greater extent than the "subjective" one.

Analysis of Data: H_0 vs. H_2 and H_3

Additional interests in the study were concerned with the association between program effectiveness and certain programer variables. Under H_2 and H_3 we expect to find a positive correlation between the effectiveness of the produced program and both the academic preparation and teaching experience of the programer.

The data obtained for this analysis, in addition to the class means of post-test scores, were self-reports by the programers of their prior training and experience. The statistic used herein to measure this association is the Spearman rank correlation coefficient, rho, here represented by r_s.

The data for the calculation of r_s is reported in Table 6 with the raw data already converted to ranks. Correlations between program effects and each of the programer variables are reported separately for each treatment group in Table 5.

Reference to a standard table of critical values of r_s shows none of the observed correlations to be reliable at a level of significance of $p \leq .05$. Thus we cannot reject H_0 that the observed values are the result of chance fluctuations.

Summary of Findings

Two samples of programers revised a preliminary "draft" version of an audio-visual instructional program by preparing a 15-minute videotaped supplementary lesson. The programers in one sample based their revisions on a subjective analysis of the program's objectives; those in the other, on test data obtained after a trial run of the original. All versions of the program and the post-test were then administered in sixth-grade classes.

The effects produced by the programs were found to occur in the predicted order of magnitude at a significance level of $p < .0013$ the original program was the least effective; the subjective revisions, as a group, were more effective; and the objective revisions, as a group, were most effective of all.

No significant correlations were found between program effectiveness and the amount of either academic training or teaching experience possessed by the programer.

TABLE 5

Summary of Correlations (r_s) Between Program Effects and Selected Programer Variables

	Class Means: Subjective (I) Revisions	Class Means: Objective (II) Revisions
Years of Education	- .11	+ .18
Years of Teaching Experience	- .50	- .25

None of the above correlations are statistically significant.

TABLE 6

Summary of Data Used for Calculation of Spearman Rank Order Correlation Coefficient, r_s

	Subjective Revision Group (I)										Objective Revision Group (II)							
Programer	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Class Mean Scores by Rank (X)	4	2	5	6	9	3	1	10	7	8	1	7	6	3	5	4	8	2
Years of Academic Training by Rank (Y)	1.5	7.5	3.5	10	3.5	9	5.5	7.5	1.5	5.5	3	1.5	7	8	5	5	5	1.5
Years of Teaching Experience by Rank (Z)	3	9.5	5	9.5	1	5	8	5	7	2	6	1.5	4	7.5	3	7.5	5	1.5

$r_{s_{X_I Y_I}} = -.11$, $r_{s_{X_{II} Y_{II}}} = +.18$, $r_{s_{X_I Z_I}} = -.50$, $r_{s_{X_{II} Z_{II}}} = -.25$

None of the r_s's observed are significant at $p \leq .05$.
H_0 is not rejected.

CONCLUSIONS AND RECOMMENDATIONS

One of the important outcomes of educational research is the evolution of a body of tested methods and procedures for developing effective instructional products and practices. The underlying principles revealed by basic research which help us to understand the instructional process are only useful to the extent that they lead directly to more effective instructional methods. Often we learn of principles and procedures which have been applied effectively by others only to find that we do not obtain the same beneficial results when we try them for ourselves.

The object of this study was to try out an experimental design for comparing alternative instructional procedures which would provide evidence that instructors other than the experimenter, by following a specific procedure, could improve the effectiveness of their instructional materials. Therefore, the design called for a number of programers to try alternative procedures to determine if one of the procedures led to the production of more effective materials than the other.

The alternative programing procedures selected for a test of this model concerned two analysis procedures in common use for analyzing and revising an instructional program: one based upon an "expert" subjective analysis of the learning objectives; the other, upon an objective analysis of test data obtained from a trial run of the existing program. It is a tradition of long standing among educational publishers and producers to seek the review of "experts" during the development of instructional products intended for publication or mass distribution. Typically, reviews by experts are conducted while the materials are under development in order that the final product may be as effective as possible. On the other hand, during the past decade, arising from techniques developed within the programed learning movement, more empirical analysis procedures have been advocated. These procedures involve testing the developmental materials with representative samples of the intended learner population and analyzing the results of tests based upon the program's learning objectives.

If we are to refine the technology of program development, if we are to develop a body of increasingly reliable methods and procedures for improving instructional effectiveness, it seems reasonable that we should seek to specify not only what must be done, but also who must do it. Little has been done, however, to determine if the ability to conduct these kinds of analyses and to use the findings effectively are common traits among people in general or if there are specific psychological or demographic characteristics associated with these abilities. It is possible, for example, that the expert review may constitute an effective developmental procedure if we become very expert in our selection of "experts." It is also possible that the analysis of data derived from pre-testing may be a more effective procedure for some programers than for others and we may profit from becoming more expert in our selection of programers.

The model tested in this study, therefore, compared the effects of two competing programing procedures upon the products of various programers who applied them. A number of programers then followed one or the other procedure in developing an audio-visual instructional program. The effectiveness of the resulting programs was then analyzed to (1) compare the effectiveness of the procedures, and (2) to relate the effectiveness of the resulting programs to certain demographic traits of the programers.

It was found that the analysis of test data derived from a tryout of the original program was reliably more effective than the analysis of learning objectives alone when a number of programers used these procedures as a basis for improving a program. However, a correlational analysis failed to show that either the amount of academic training or the amount of prior teaching experience (both commonly assumed to be related to instructional effectiveness) were reliably related to the effectiveness of the programs.

It should be noted, however, that the *negative* correlation found between the effectiveness of programs revised subjectively and the amount of the programers' teaching experience was of such magnitude as would have been reliable at a one-tail level of significance of $.05 < p < .10$ had this been the predicted direction of the relationship. An outcome of this sort would tend to lend support to the findings of earlier research (see Rothkopf, 23, and Merrill and McAshan, 20) which showed high *negative* correlations between expert predictions and objective measures of program effectiveness. If it should turn out that the subjective analysis of experts consistently leads to erroneous predictions, it is possible that revisions based upon such predictions may actually diminish the effects of the original program. While this outcome in the present study can be explained as a chance fluctuation, both the magnitude and direction of the correlation surely raise questions worthy of further investigation.

It would appear then that the experimental model employed in the study offers a viable procedure for assessing the relative effectiveness of alternative programing procedures as they are applied by a sample of programers. The analysis of student test data has been widely advocated as an effective way to revise an instructional sequence and the results of this study tend to support this claim. We are able to conclude that a population of state-credentialed teachers, at least, seems reliably able to make effective use of objective test information to improve their instruction.

Clearly, similar experiments might be conducted to test the effects of other programing procedures. Competing theoretical learning principles, for example, might be translated into competing sets of programing procedures which could then be tested in this manner. The results of such experiments would contribute to the development of a technology of education in two ways: first, by providing empirical support for the theoretical position and, second, by specifying a reliable procedure for applying laboratory theory to a practical teaching problem.

One of the factors inhibiting the development of a reliable technology of instruction concerns our inability at this time to clearly describe whether the effects of a procedure justify its cost. While most current research, as Lumsdaine and others have repeatedly pointed out (4, 15, 16, 19), reports the reliability of the observed effects, very little reports upon the amount of the effect in terms which are useful for making decisions to use or not to use the procedure which produced it. Markle (18), too, has pointed out that the typical developmental study fails to generate data regarding the cost-effectiveness of the procedures used as they might compare to some alternative procedures. Any attempt to assess the cost-effectiveness of programing procedures leads to the inevitable problem of quantifying both the learning benefits and the costs of achieving them. At present the parameters of these variables are too little known to develop a satisfying cost-benefit model which allows a rational decision-making analysis in educational contexts.

In training contexts, on the other hand, (as the distinction is commonly employed) we are provided classic models for cost-benefit analysis. Markle, for example, may not have provided an experimental comparison of his procedures with some alternative procedures; nevertheless, the cost-benefit of his effort may be fully assessed. His objective was to realize student-time savings against a standard test of specific acquired knowledge and skill in a context where student-time was valued in dollars. The cost of the effort may be traded off directly against the value of student-time to determine the cost-benefit derived. In the educational context, however, the issues are not nearly so clearly defined. Student-time is not valued in dollars, nor are there widely accepted standards for valuing student learnings. In the present study, for example, we were able to show only that with additional efforts we reliably achieved additional effects; moreover, the procedure which involved the most additional effort reliably produced the most additional effect. Whether the effects of the procedures are worth the efforts needed to bring them about cannot as yet be answered. How, then, are we to develop appropriate cost-benefit models for the instructional procedures we employ in educational contexts?

One model that might be suggested would describe learning effects as a function of costs, rather than as a function of learner time. After all, school operating costs are valued in dollars which provides at least one of the relevant parameters for a satisfying cost-benefit analysis. In any case, if there is to develop an effective educational technology concerned with applying knowledge of the basic sciences to the practical arts of teaching, educational research models are needed which address themselves to the family of problems associated with teaching more students more effectively while holding the cost-basis relatively constant. While this problem was not dealt with directly in the foregoing experiment, it is worthy of note that the proposed model in addition to testing the relative effectiveness of two procedures would additionally provide a basis for analyzing the relative cost-benefits.

REFERENCES

1. American Educational Research Association, American Psychological Association, and Department of Audio-Visual Instruction (National Education Association) Joint Committee on Programmed Instruction and Teaching Machines. "Criteria for Assessing Programmed Instructional Material." Audiovisual Instruction 8: 84-88; 1963.

2. American Institutes for Research. English Money: An Auto-Instructional Program. Pittsburgh: The Institute, 1961.

3. Edling, J.V. A Study of the Effectiveness of Audio-Visual Teaching Materials When Prepared According to the Principle of Motivational Research. Monmouth: Oregon State System of Higher Education, NDEA Project No. 221, 1963.

4. Gropper, G.L., Lumsdaine, A.A. and Shipman, V. Improvement of Televised Instruction Based on Student Responses to Achievement Tests. Pittsburgh: American Institutes for Research, Report No. 1, Studies in Televised Instruction, 1961.

5. Gropper, G.L. and Lumsdaine, A.A. The Use of Student Response to Improve Televised Instruction: An Overview. Pittsburgh: American Institutes for Research, Report No. 7, Studies in Televised Instruction, 1961.

6. Holland, J.G. "Teaching Machines: An Application of Principles From the Laboratory." In Lumsdaine, A.A. and R. Glaser, eds., Teaching Machines and Programmed Learning: A Source Book. Washington, D.C.: National Education Association, 1960. Pp. 215-228.

7. Hovland, C.I., Lumsdaine, A.A. and Sheffield, F.D. Experiments on Mass Communication. Princeton: Princeton University Press, 1949. Chapter 14.

8. Jonckheere, A.R. "A Distribution-Free k-Sample Test Against Ordered Alternatives." Biometrika, 41: 133-145; 1954.

9. Kendall, M.G. Rank Correlation Methods. London: Griffin, 1948.

10. Levonian, E. "Development of an Audience-Tailored Film." Audiovisual Communication Review, 8: 62-68; 1960.

11. Levonian, E. "Opinion Change as Mediated by an Audience-Tailored Film." Audiovisual Communication Review, 11: 104-113; 1963.

12. Lindquist, E.F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Co., 1953. Chapter 7.

13. Lumsdaine, A.A. "Instruments and Media of Instruction." In Gage, N.L. ed., Handbook of Research on Teaching. Chicago: Rand-McNally, 1963. Chapter 12.

14. Lumsdaine, A.A. "Educational Technology, Programed Learning and Instructional Science." In National Society for the Study of Education 63rd Yearbook, Theories of Learning and Instruction. Chicago: University of Chicago Press, 1964. Chapter 16.

15. Lumsdaine, A.A. and May, M.A. "Mass Communication and Educational Media." In Annual Review of Psychology. Palo Alto: Annual Reviews, 1965. Pp. 475-534.

16. Lumsdaine, A.A. "Assessing the Effectiveness of Instructional Programs." In Glaser, Robert, ed., Teaching Machines and Programed Learning, II. Washington, D.C.: National Education Association, Department of Audiovisual Instruction, 1965. pp. 267-320.

17. Markle, David G. "Empirical Film Development." Journal of the National Society for Programed Instruction, IV: No. 6; July 1965.

18. Markle, David G. The Development of the Bell System First Aid and Personal Safety Course. Palo Alto: American Institutes for Research, 1967.

19. May, M.A. and Lumsdaine, A.A. Learning from Films. New Haven: Yale University Press, 1958. Chapter 2.

20. Merrill, I.R. and McAshan, H.H. "Predicting Learning, Attitude Shift, and Skill Improvement from a Traffic Safety Film." Audiovisual Communication Review, 8: 263-74; 1960.

21. Meyers, Lawrence, Jr. An Experimental Study of the Influence of the Experienced Teacher on Television. Syracuse: Syracuse University, 1961.

22. Orlando, Robert. Analyzing System Components and Functions for Research and Development in Programmed Instruction. Paper presented to the National Society for Programed Instruction, April, 1964. (Ditto.)

23. Rothkopf, E.Z. Some Observations on Predicted Instructional Effectiveness by Simple Inspection. New Jersey: Bell System Laboratories, 1963. (Offset.)

24. Sheffield, F.D. "The Effects of Presenting 'One Side' versus 'Both Sides' in Changing Opinions on a Controversial Subject." In Hovland, C.I., et al, Experiments on Mass Communication. Princeton: Princeton University Press. 1949. Chapter 8.

25. Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

26. Silberman, H., et al. Use of Exploratory Research and Individual Tutoring Techniques for the Development of Programming Methods and Theory. Santa Monica: System Development Corporation, 1964.

27. Twyford, L. "Film Profiles." Instructional Films Research Reports (SDC 269-7-23). Port Washington: U.S. Naval Special Devices Center, 1951. Unpublished doctoral dissertation, Pennsylvania State University, 1950.

28. VanderMeer, A.W., H.E. Thorne and R. Montgomery. "An Investigation of the Improvement of Informational Filmstrips and the Derivation of Principles Relating to the Effectiveness of these Media: Studies I--III." Abstracted in VanderMeer, A.W. "Ten Recent Title VII Reports: A Perspective." Audiovisual Communication Review, 9: No. 6; 1961.

29. VanderMeer, A.W., and Philip Smith. An Investigation of the Improvement of Educational Motion Pictures and a Derivation of Principles Relating to the Effectiveness of these Media. University Park: College of Education, Pennsylvania State University, April, 1965. (NDEA Title VII Project No. 225.)

30. Zuckerman, J.V. "Predicting Film Learning by Pre-release Testing." Audiovisual Communication Review, 2: 49-56; 1954.

APPENDIX A

30-ITEM TEST USED FOR ITEM ANALYSIS

AUDIO SCRIPT TO ACCOMPANY TRYOUT TEST

36. In words, write the total value of these coins.

37. In words, write the total value of the coins named here.

38. In symbols, write the expression for two pounds, six shillings, fourpence, hay-p'ny. Listen again. Two pounds, six shillings, fourpence, hay-p'ny.

39. In words, write the value of a florin.

40. With these coins you pay two-and-nine. The clerk hands you one coin in change. What is the name of that coin?

41. To pay one-and-three you hand the clerk a half-crown. In words, write the value of your change.

42. You owe a friend six shillings ninepence. You give him these coins as part payment. What single coin is still needed to pay him in full?

43. From the total amount shown here, you spend the coins shown here. In symbols write the amount you have left.

44. You have four shillings sixpence in your pocket and you spend half-a-crown. What is the name of the single coin you have left?

45. From the amount shown here you spend a florin and a thrup'ny bit. Name the two coins you still have left.

46. From the amount shown here you spend the coins shown below. In symbols, write the amount you have left.

47. You have ten shillings. You spend six shillings, thrup'nce, hay-p'ny. In symbols, write the amount you have left.

48. You have one pound. You spend the amount shown here. In symbols, write the amount you have left.

49. From the amount shown above, you spend the amount shown below. Name the two coins you still have left.

50. From the amount shown above, you spend the amount shown below. Name the three coins you still have left.

36. Write the names of the following coins: a thrup'ny bit, tuppence, and a hay-p'ny.

37. In symbols, write the total value of the coins shown.

38. In symbols, write the amount shown here.

39. If one-and-six were taken from the coin shown here at the left, which single coin would you get in change, A, B, C, or D?

40. The amount shown is taken from the coins shown at the left. In symbols, write the amount you should receive in change.

41. To pay the amount shown, you give a florin and a shilling. Which coins, A, B, C, or D are your correct change?

42. You give the coins shown to pay the amount shown. In words, write the value of your change.

43. You give the coins shown to pay the amount shown. In words, write the value of your change.

44. You owe a friend the amount shown. As part payment, you give the coins shown. Which coins, A, B, C, or D do you still owe?

45. You have six shilling. You spend half-a-crown. Which coins, A, B, C, or D do you still have left?

46. You have the amount shown, and you spend a florin and a thrup'ny bit. In symbols, write the amount you still have left.

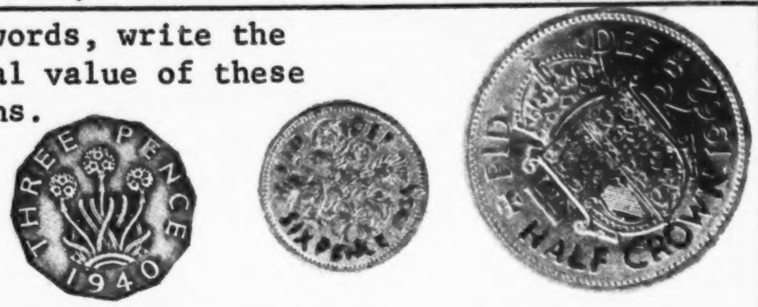
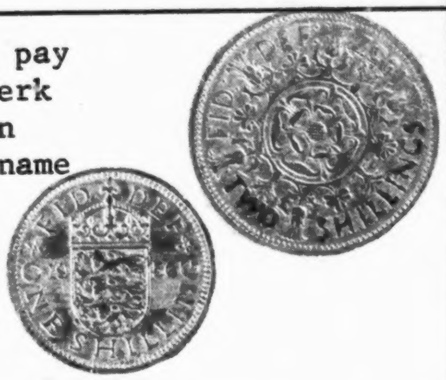
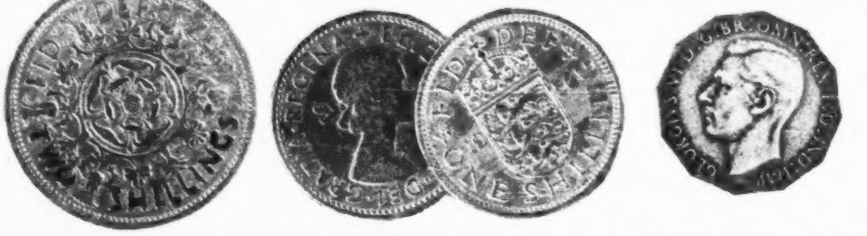
47. You give a half-crown and a shilling to pay a bill for two-and-nine. In symbols, write the amount of change you get.

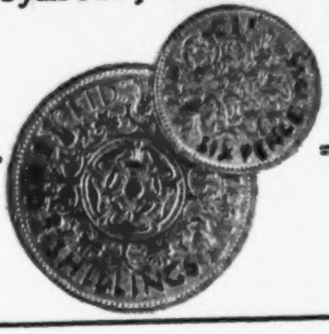
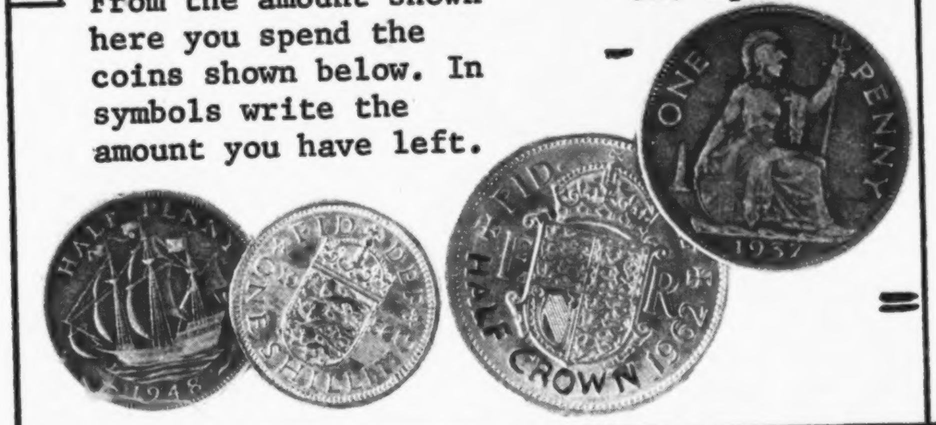
48. From the amount shown here, you spend the coins shown here. In words, write the amount you have left.

49. From the amount shown above, you spend the amount shown below. Which of the following coins, A, B, C, or D, do you have left?

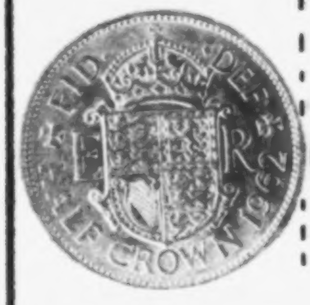
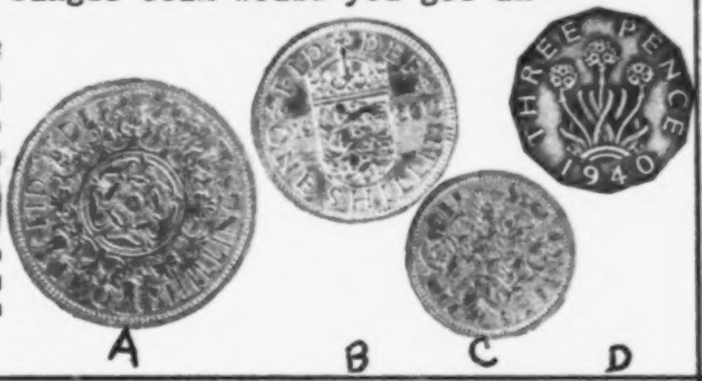
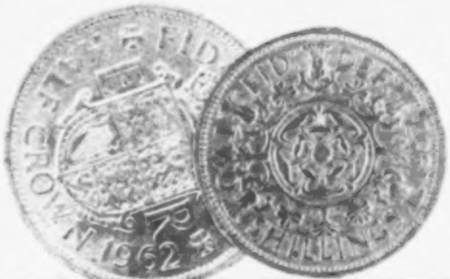
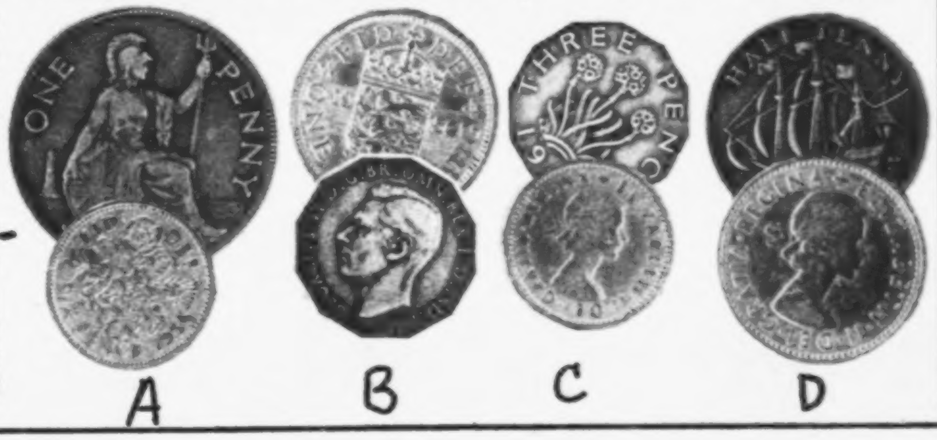
50. From the amount shown above, you spend the amount shown below. In symbols, write the amount you have left.

ENGLISH MONEY TEST

QUESTION	ANSWER
36 In words, write the total value of these coins. 	
37 In words, write the total value of the coins named here. 1 HALF CROWN + 1 FLORIN + 1 THREEPENNY BIT =	
38 ?	
39 In words, write the value of a FLORIN.	
40 With these coins you pay two-and-nine. The clerk hands you one coin in change. What is the name of that coin? 	
41 To pay one-and-three you hand the clerk a HALF CROWN. In words, write the value of your change.	
42 You owe a friend six shillings, nine pence. You give him these coins as part payment.  What single coin is still needed to pay him in full?	

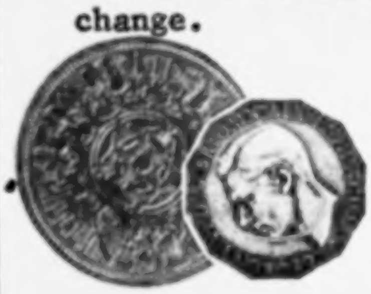
43 From the total amount shown here, you spend the coins shown here. In symbols, write the amount you have left. 14 SHILLINGS, NINE PENCE -  TWO SHILLINGS SIX PENCE =	
44 You have four shillings, six pence in your pocket and you spend half-a-crown. What is the name of the single coin you have left?	
45 ?	
46 From the amount shown here you spend the coins shown below. In symbols write the amount you have left. 9s. 3½d. -  HALF PENNY 1948 SHILLING HALF CROWN 1962 ONE PENNY 1937 =	
47 ?	
48 ?	
49 ?	
50 ?	

ENGLISH MONEY TEST II

QUESTION	ANSWER
36 ?	
37 In symbols, write the total value of the coins shown. 6 HALFCROWNS + 3 FLORINS + 2 SHILLINGS + 6 PENCE =	
38 In symbols, write the amount shown here. 8 shillings, 2 pence, halfpenny =	
39 If one-and-six were taken from the coin shown here, which single coin would you get in change?   	
40 The amount shown is taken from the coins shown at the left. In symbols, write the amount you should receive in change.  - three-and-three =	
41 To pay the amount shown you give a FLORIN and a SHILLING. Which coins are your correct change? 2/3-  	

42

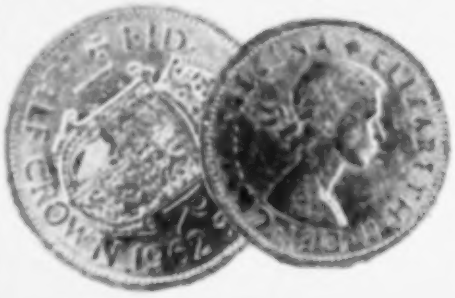
You give the coins shown to pay the amount shown. In words, write the value of your change.



- 2 shillings, halfpenny =

43

You give the coins shown to pay the amount shown. In words, write the value of your change.

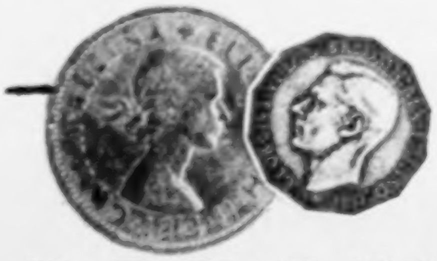


- 2/11 =

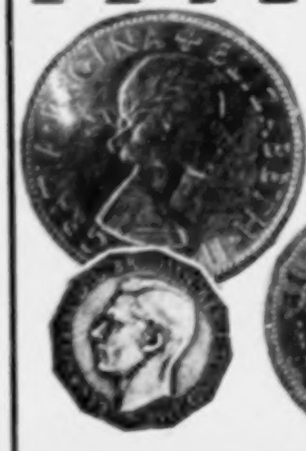
44

You owe a friend the amount shown. As part payment you give the coins shown. Which coins do you still owe?

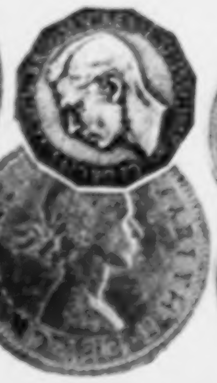
5/- –



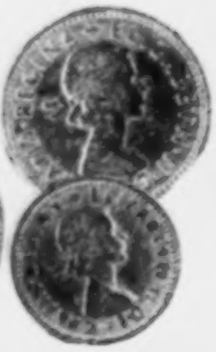
=



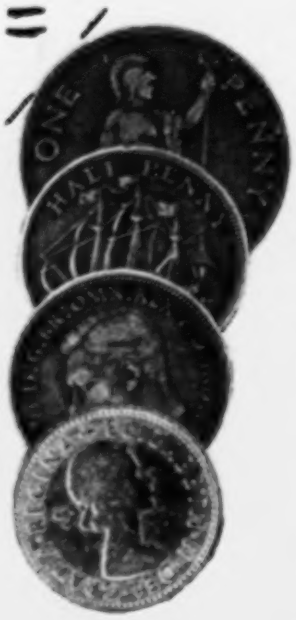
A



B

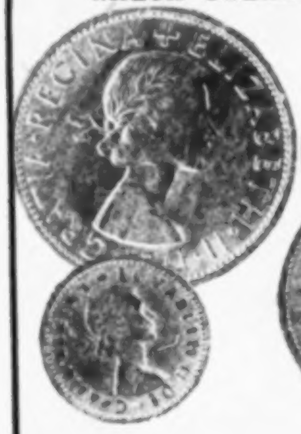


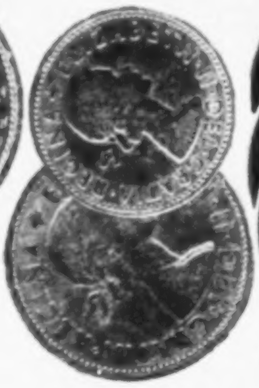
C

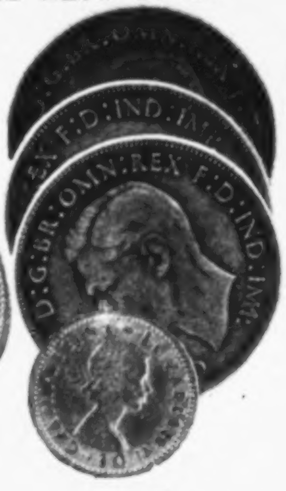


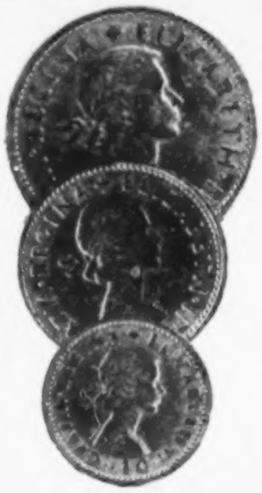
D

45 You have six shillings. You spend half-a-crown. Which coins do you still have left?


A


B


C

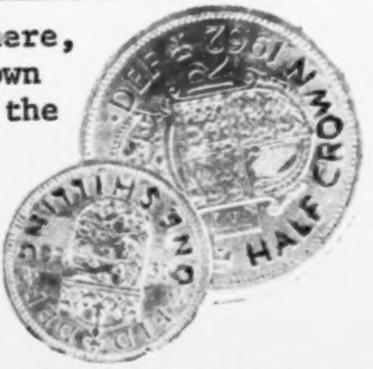

D

46 ?

47 You give a HALF CROWN and a SHILLING to pay a bill for two-and-nine. In symbols, write the amount of change you get.

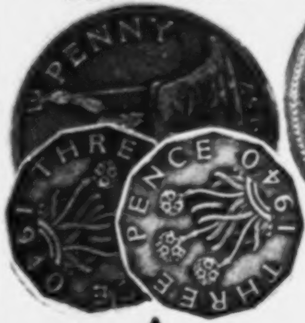
48 From the amount shown here, you spend the coins shown here. In words, write the amount you have left.

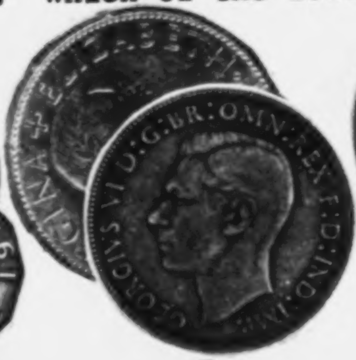
6/11 -

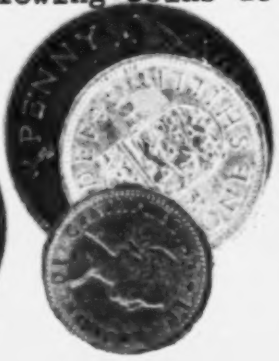

DEF 1952
HALF CROWN
ONE SHILLING

=

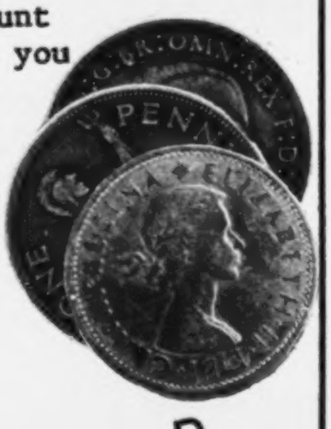
49 From the amount shown above you spend the amount shown below. Which of the following coins do you have left?


PENNY
THREE PENCE 1940
A


B



C



D

50 ?

APPENDIX B

INSTRUCTION BOOKLET

APPENDIX B: INSTRUCTION BOOKLET

You are going to plan a 15-minute lesson on the subject of English Money for sixth graders to supplement an existing slide-tape program. The existing program, plus your videotaped supplement, will be presented to some sixth-grade classes in the area. A number of teachers are working on this same project. Your job, through your videotaped lesson, is to teach the children what you think they need to learn to pass an achievement test on the subject immediately following the lesson. We are interested to learn which teaching strategy is most effective as measured by the test scores following the lesson.

In this booklet you will find the following materials to help you in this task:

1. A general statement of the learning tasks;
2. A description of the stimulus-response elements as they are managed within the slide-tape program;
3. A "test-item generator" which describes the kinds of test items which are likely to appear in an achievement test on the subject;
4. A script of the slide-tape program for your reference and review;
5. (Treatment II only) A copy of the final achievement test showing the proportion of sixth graders who answered each item correctly after a trial run of the slide-tape program.

You will have a maximum of six hours to complete your tape from the time you start. You should allow no more than three hours for preliminary planning in order to provide sufficient time for rehearsing and recording portions of the lesson in the studio. Our staff will be in the area at all times to assist you in using the equipment, rehearsing the lesson, or to answer questions of a technical nature.

Good Luck.

General Statement of the Learning Tasks

The student encountering English Money for the first time is generally exposed to the actual coins and bills in circulation, to price tags and labels, and to merchants who quote prices and label pieces of money in colloquial speech. The tasks the student is called upon to perform in the presence of these aural and visual symbols involve computing equivalences and differences within the English money system and expressing the result in terms of several symbol systems interchangeably. A typical purchase might involve a merchant quoting a price in colloquial speech, the selection of an appropriate set of coins by the student to cover the purchase, the computation of the correct value of change to be received, and the assessment of equivalence between the value of the expected change and the value of the change received.

The program proposes to expose the student to the expected range of stimulus classes asking him to make the typical responses he would be expected to make under "realistic" circumstances. Thus, in the presence of a variety of stimulus situations the student will be asked to express equivalences and differences in value in any of the conventional symbol systems.

For administrative purposes, spoken responses are excluded from the program.

General Objectives Established for the Program

The following general objectives are established for the program:

1. Recognition and labeling of English coins in common circulation.

2. Computing equivalences and differences of value interchangeably among coins, written and spoken verbal statements, and numerical expressions according to the common practice.

3. Expressing values interchangeably in terms of coins, written verbal statements, and numerical expressions according to the common practice.

General Objectives NOT Established for the Program

The following statements describe certain general objectives specifically excluded from those established for the program:

1. Computing American monetary equivalents of British units.

2. Estimating costs of common items in British commerce.

3. Spoken responses of any sort on the part of the student.

4. S-R combinations involving coin-bill configurations valued in excess of £1.

Stimulus Classes

The student will be presented photographic representations of English coins and their respective labels as follows: half-penny, three-penny bit, shilling, florin, and half-crown. He will be presented a range of values from 1/2 d. to several pounds expressed in terms of written and spoken words, and numerical expressions; from 1/2 d. to 19/11 1/2 expressed in terms of coins as well.

Response Classes

The student will be asked to respond to these stimuli by expressing equivalences and differences in value. These responses are to take the form of written verbal statements, numerical expressions, and selected photographs showing coin configurations. For ease of administration, spoken responses are excluded from the program.

The following table defines the stimulus-response elements as they are managed within the program.

Program Element	As Stimulus	As Response
Coin Configuration	A photograph of coins is used as a stimulus pattern.	S selects from several alternative photos of coins to express equivalence or difference.
Coin Labels, spoken	The names of coins are spoken colloquially to S (eg. "Thrup'ny Bit").	S speaks the names of coins colloquially. However, spoken responses are excluded from this program.
Coin Labels, written	The names of coins are written out correctly spelled (ie. not as colloquially pronounced, eg. "Three-Penny Bit").	S writes the names of coins as they would be correctly spelled (ie. not as colloquially pronounced, though spelling is NOT an objective of the program).
Value Statements, spoken	A value is spoken to S colloquially (eg. "Tuppence Hayp'ny").	S expresses a value colloquially. However, spoken responses are excluded from this program.
Value Statements, written	A value is written in words (eg. "Two Pence Half-Penny").	S expresses a value in words by writing them as a response (eg. "Two Shillings").
Value Expressions	A value is expressed in terms of numerical symbols according to the conventional practice (eg. "12s.1d=12/1").	S expresses a value in numerical symbols according to the conventional practice (eg. "6/5=6s.5d.").

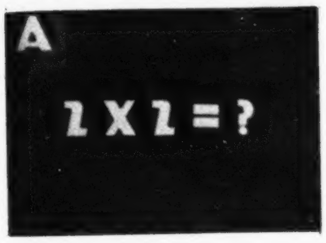
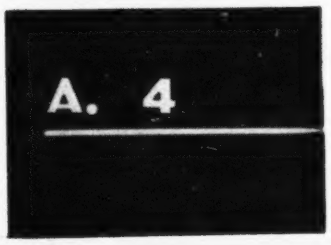
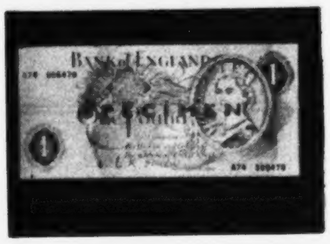
BEHAVIORAL OBJECTIVES EXPRESSED
AS TEST-GENERATING STATEMENTS

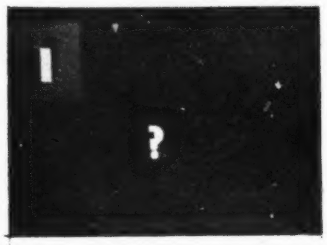
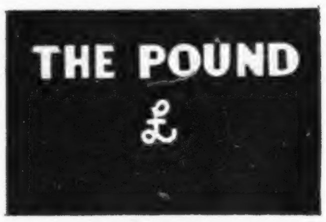
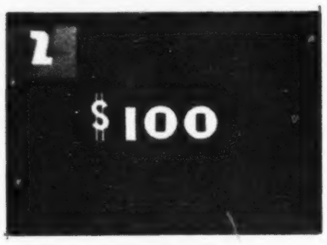
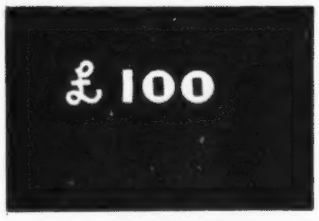
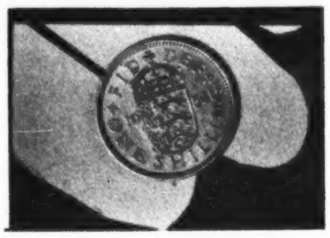
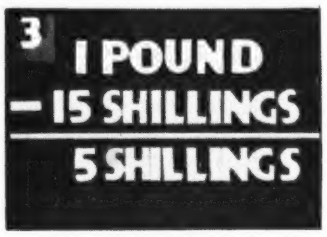
S is given any of the following:	And is asked to subtract any of the following:	And to express his answer by any of the following:
- The names of some coins (written out).	- The value of some coins named in writing.	- Naming a correct set of coins in writing.
- The names of some coins (spoken aloud).	- A given amount of money spoken aloud.	- Stating in words (written) a correct amount of money.
- A set of coins (in photographic form).	- A given amount of money written out in words.	- Expressing a correct amount of money by writing a conventional numerical expression.
- An amount of money (written as a conventional numerical expression).	- An amount of money written as a conventional numerical expression.	- Selecting a correct set of coins from several sets shown in photographs.
- An amount of money (written in words).	- The value of a given set of coins (in photographic form).	(Oral responses excluded)
- An amount of money (spoken aloud).	- The value of some coins named aloud.	

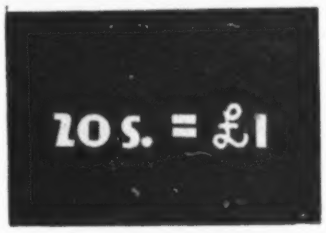
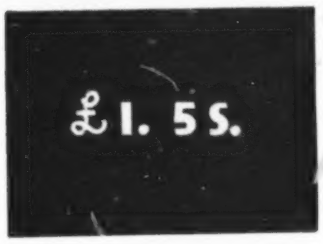
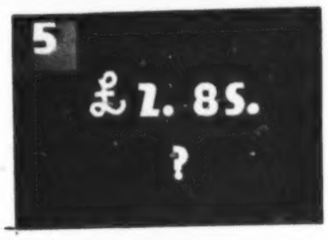
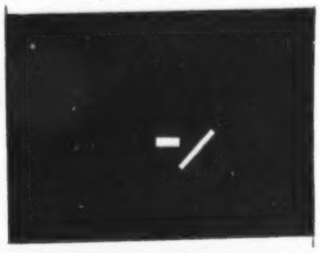
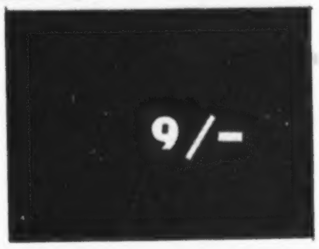
SCRIPT OF BASE PROGRAM

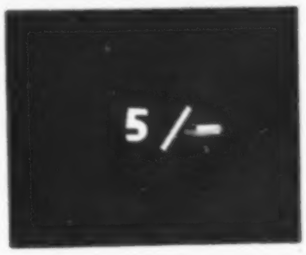
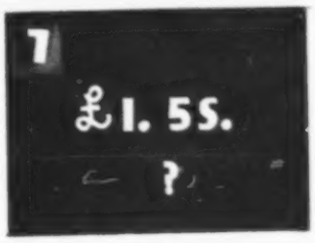
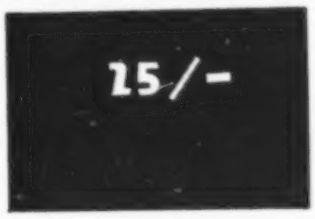
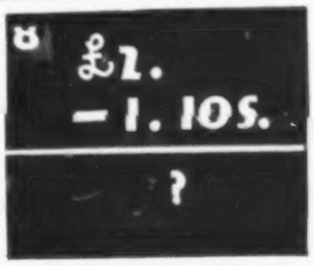
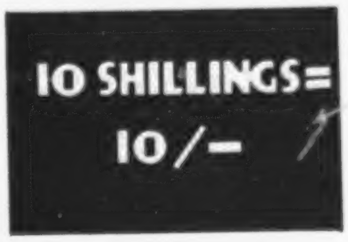
ENGLISH MONEY

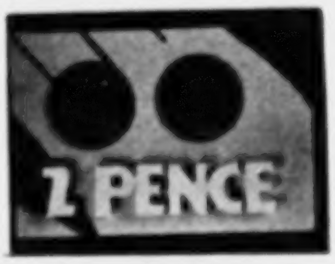
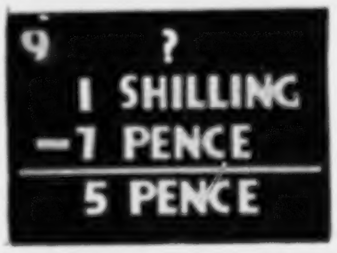
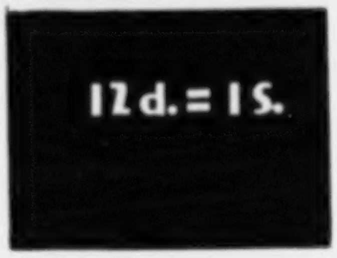
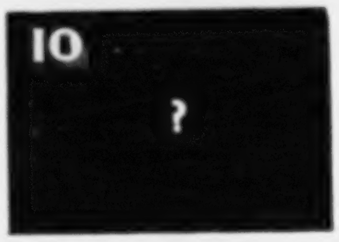
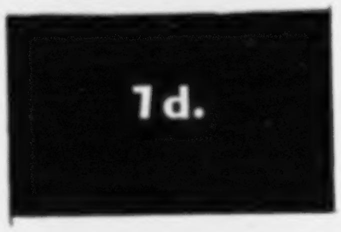
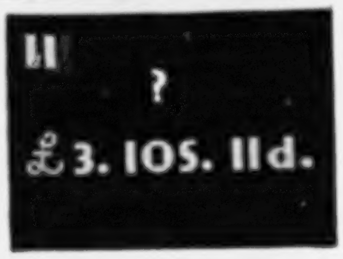
A Fixed-Pace Audio-Visual Program

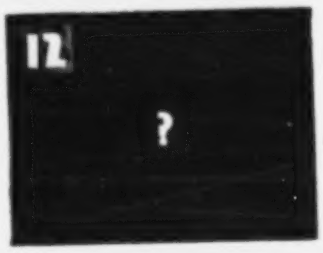
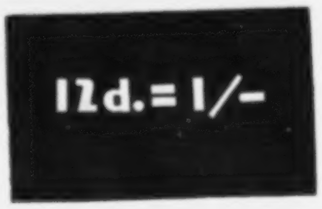
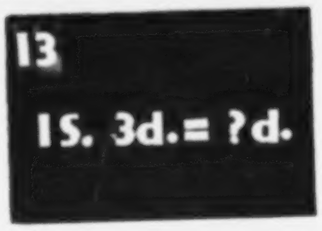
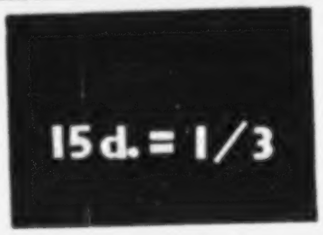
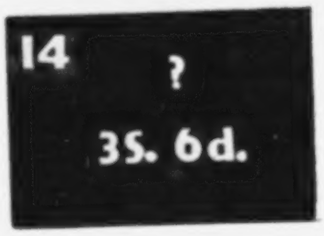
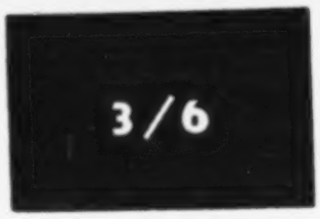
VISUAL	AURAL
 ENGLISH MONEY	This is a lesson about English money. To teach you about English money I am going to show you some slides as I talk to you. Also, I will ask you some questions as we go along. The question slides will look like this.
 A 2 X 2 = ?	Notice the figure in the upper-left corner. Whenever you see a question slide, write your answer in the correct space on your answer sheet. Try to answer every question. It will help you learn. Answer this question in the "A" space on your answer sheet.
 A. 4	You should have put a "4" in answer space "A" on your answer sheet. If you have any questions, raise your hand now. (STOP TAPE -- THEN START) Now let's see how much you can learn about English money.
 ENGLISH MONEY	English money seems strange at first because it is not based upon the decimal system.
	The basic unit of English money is the Pound.

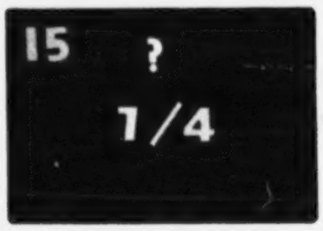
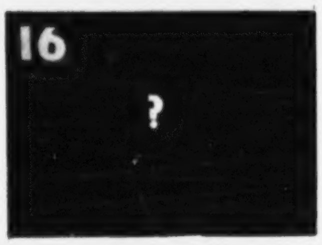
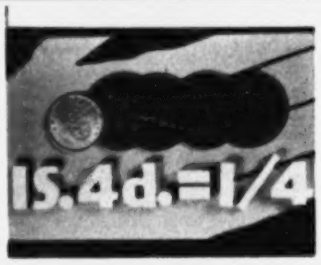
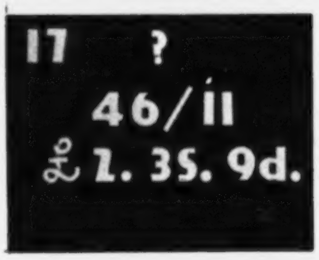
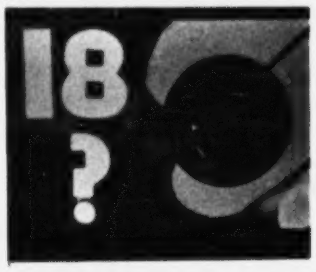
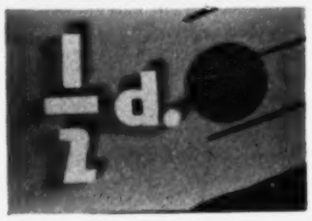
 1 ?	Here is your first question. Write your answer in space number 1 on your answer sheet. What is the basic unit of English money?
 THE POUND £	The pound is the basic unit of English money. The symbol shown here is the symbol for pound. Draw this symbol three times in the scratch area of your answer sheet.
 2 $100	This is how we write the symbol for one hundred dollars. Now for question number two, How would you write the symbol for one hundred pounds?
 £100	This is how to write the symbol for one hundred pounds. Notice that the pound sign comes before the amount.
	This coin is called a shilling.
 3 1 POUND – 15 SHILLINGS 5 SHILLINGS	If you gave a one pound note for something costing fifteen shillings, you would get five shillings in change. How many shillings must there be in one pound?

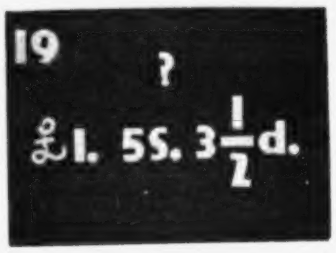
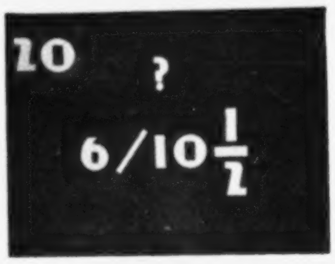
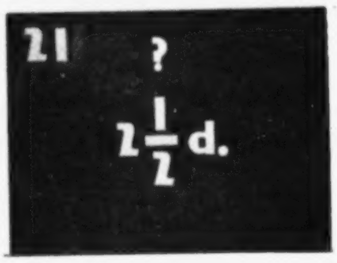
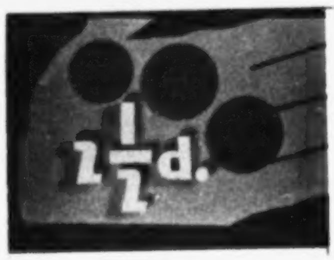
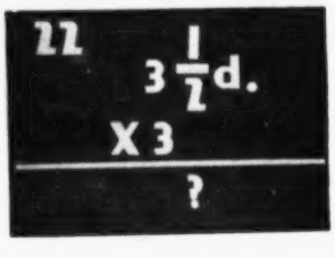
 20s. = £1	There are twenty shillings to the pound. The symbol for shillings is the small "s" placed after the amount.
 4 ?	How would you write the symbols for one pound five shillings?
 £1. 5s.	This is how one pound five shillings is written in symbols. notice that a period is placed after each denomination.
 5 £2. 8s. ?	How would you write in words the amount shown here? (PAUSE) You would write this as two pounds eight shillings. Sometimes a price is shown in shillings alone.
 -/	Whenever a price is shown in shillings alone, you will see this symbol. Draw this symbol three times in the scratch area of your answer sheet.
 9/-	The symbol shown here is read nine shillings.
 6 ?	How would you write the symbols for five shillings alone?

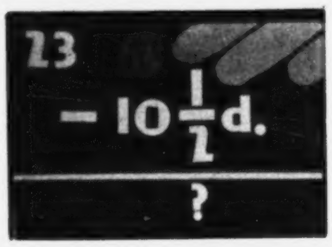
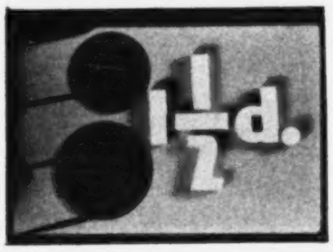
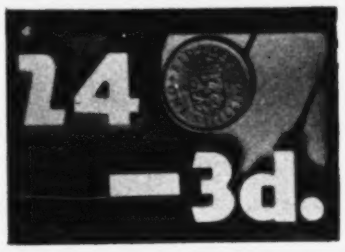
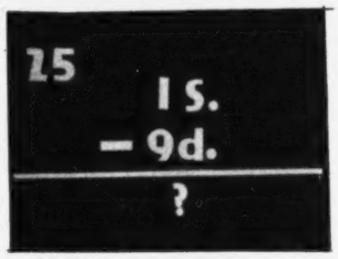
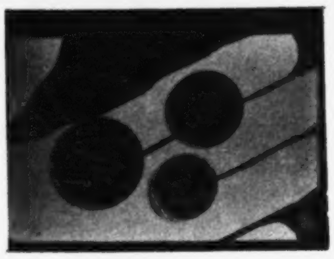
 5/-	This symbol shows five shillings alone.
 7 £1. 5S. ?	How would you write the symbols for this amount in shillings alone?
 25/-	Twenty-five shillings. Remember, there are twenty shillings to the pound. So twenty shillings plus five shillings gave us the twenty-five shillings shown here.
 8 £2. − 1. 10S. ?	If you gave two pounds for something costing one pound ten shillings, how much change would you receive? Write *your answer* first in words, then in symbols.
 10 SHILLINGS= 10/-	You would receive ten shillings in change.
	This coin is a penny.

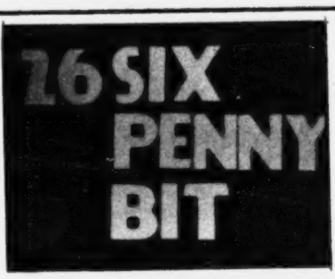
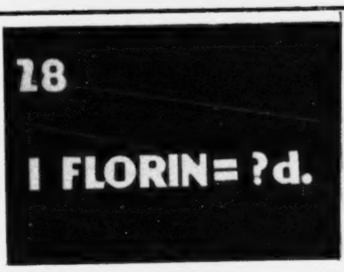
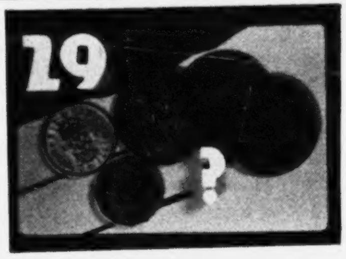
 	The plural of penny is pence. For example, you can buy a packet of gum for two pence.
 9 ? 1 SHILLING −7 PENCE 5 PENCE	If you gave one shilling for something costing seven pence, you would receive five pence in change. How many pence must there be in a shilling?
 12 d. = 1 S.	There are twelve pence in one shilling. Note that pence are expressed by the symbol "d".
 10 ?	How would you write the symbol for seven pence?
 7 d.	This is the symbol for seven pence.
 11 ? £3. 10S. 11d.	How do you express in words the amount shown here? (PAUSE) This amount is expressed as three pounds ten shillings eleven pence.

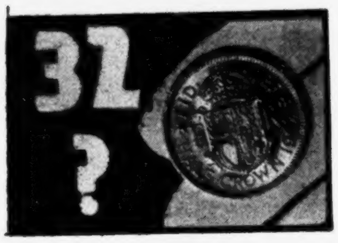
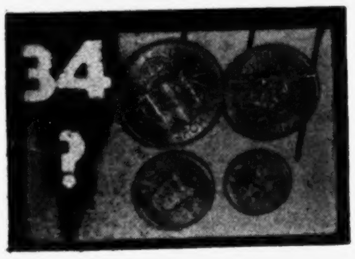
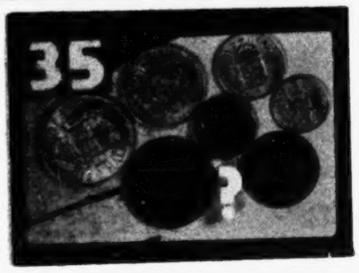
 12 ?	How many pence are in a shilling?
 12d.= 1/-	There are twelve pence in one shilling.
 13 1S. 3d.= ?d.	You could pay for a magazine priced at one shilling three pence with pence alone. How many pence would you need?
 15d.= 1/3	You would need fifteen pence. Notice how the slash is used here to show both shillings and pence.
 14 ? 3S. 6d.	How would you use the slash symbol to write the amount shown here?
 3/6	This symbol means three shillings sixpence. In England it is usual to omit the words shilling and pence when speaking of sums of money. So this symbol is read as three and six.

	Write how you would read the symbol shown here and what it means? (PAUSE) The symbol is read as seven and four. It means seven shillings fourpence.
	If a shopkeeper told you that a pound of apples cost one and four, how much would you pay him for the apples?
	You would pay him one shilling fourpence.
 17 ? 46/11 £ 2. 3S. 9d.	If one pair of slippers was marked forty-six and eleven and another was marked two pounds three shillings ninepence, which pair would cost the least? (PAUSE) Two pounds three shillings ninepence is less than forty-six and eleven.
	As its name suggests, this coin is worth half a penny. The English pronounce the name of this coin the hay-p'ny. Altogether, how many hay-pence are there in a shilling? (PAUSE) There are twenty-four hay-pence in a shilling.
	This is the way to write the symbol for a hay-p'ny.

	This symbol would be read as six pence hay-p'ny.
	In words, write the amount shown here. (PAUSE) In words, this amount is read as one pound, five shillings, three pence, hay-p'ny.
	In words, write the amount shown here. (PAUSE) This amount is read as six shillings, tenpence, hay-p'ny.
	In England, if you bought something for tuppence, hay-p'ny, shown here, you could give the clerk three coins and not wait for change. What three coins make tuppence, hay-p'ny?
	Two pence plus one hay-p'ny make tuppence hay-p'ny.
	One postcard costs thrupence hay-p'ny. For three postcards, how much would you have to pay? (PAUSE) You would pay tenpence hay-p'ny.

 	If you gave one shilling for something costing tenpence hay-p'ny, how much change should you get?
 	You should get penny hay-p'ny in change.
  THREE PENNY BIT	In England this coin is called a thrup'ny bit. It is worth thrup'nce.
 	In London you could buy a newspaper for thrup'nce and pay the newsboy with one thrup'ny bit. But you could give him a shilling and get change. How much change should you get from a shilling? (PAUSE) You should get ninepence in change.
 	If you give a shilling for a pint of milk costing ninepence, what single coin is your correct change? (PAUSE) A thrup'ny bit would be correct change.
	These three coins of lowest denomination are the hay-p'ny, the penny, and the thrup'ny bit. In England they are called coppers.

 26 SIX PENNY BIT	This small coin is a sixpenny bit. It is the lowest denomination silver coin. If it is worth sixpence, how many sixpenny bits are in a shilling? (PAUSE) There are two sixpenny bits in a shilling.
 27 2/-	This silver florin is often called a two shilling piece. How many florins are in a pound? (PAUSE) Ten florins make a pound. Remember, there are twenty shillings to a pound.
 28 1 FLORIN= ?d.	How many pence are there in a florin? (PAUSE) There are twenty-four pence in a florin.
 29 ?	These coins do not quite add up to a florin. What single coin could you add to make up the correct amount? (PAUSE) An additional sixpenny bit is needed to make two shillings.
 HALF CROWN= 2/6	This coin is a half crown. It is the largest silver coin in circulation. A half crown is worth two shillings sixpence, or two-and-six.
 30 ?	These coins do not quite add up to a half crown. What single coin could you add to make up the correct amount? (PAUSE) An additional shilling is needed to make two-and-six.

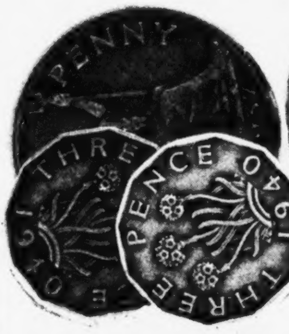
	These coins do not quite add up to a half crown. How many thrup'ny bits are needed to make up the correct amount? (PAUSE) Two thrup'ny bits are needed to make two shillings sixpence.
	You ask for change of a half crown in sixpenny bits. How many sixpenny bits will you get? (PAUSE) You will get five sixpenny bits for your half crown.
	You give these two coins for a ticket costing two-and-six. How much change should you get? (PAUSE) You should get sixpence in change.
	These are the four silver coins of England. Added together, how much are they worth? (PAUSE) Altogether they are worth six shillings.
	How much is this pocketful of change worth? (PAUSE) Altogether this change is worth six shillings fourpence ha'p'ny. These are all the coins now in circulation. This is the end of the picture lesson.

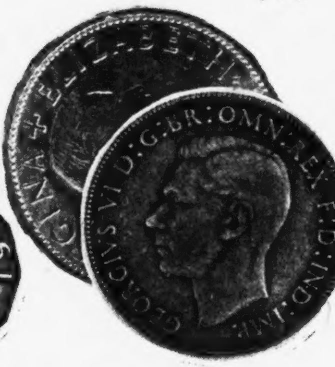
ENGLISH MONEY TEST

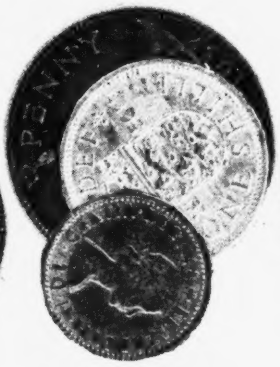
1 From the amount shown above, you spend the amount shown below. Which of the following sets of coins do you have left? A, B, C, or D?

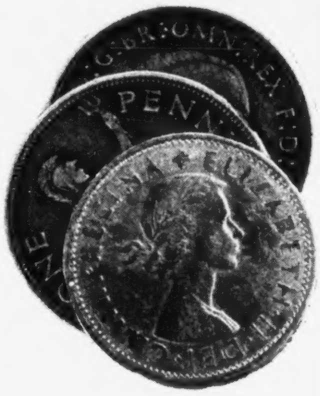
5 shillings, 4 pence
- 3 shillings, 9 pence

???

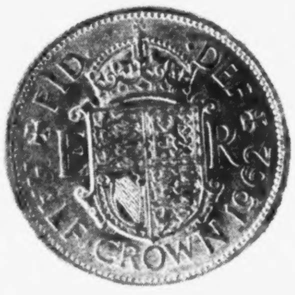

A

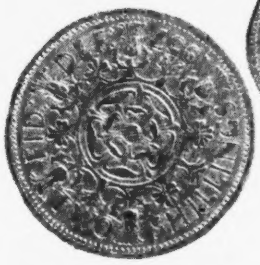

B

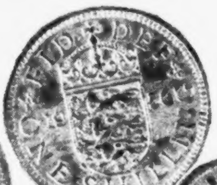

C

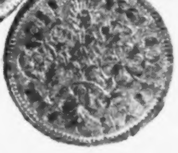

D

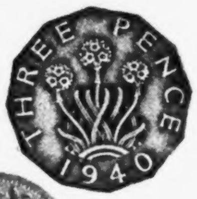
2 If one-and-six were taken from the coin shown here at the left, which single coin would you get in change? A, B, C, or D?




A

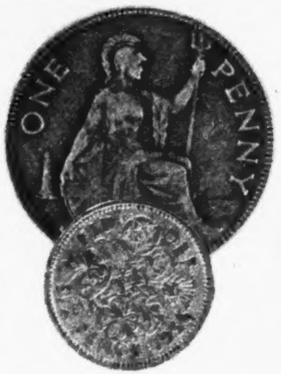

B

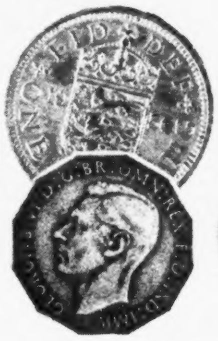

C

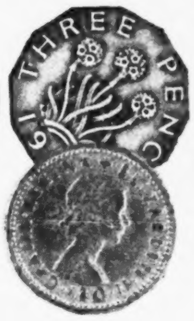

D

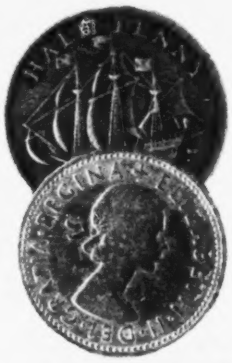
3 To pay the amount shown here at the left, you give a FLORIN and a SHILLING. Which set of coins is your correct change? A, B, C, or D?

2/3


A


B


C

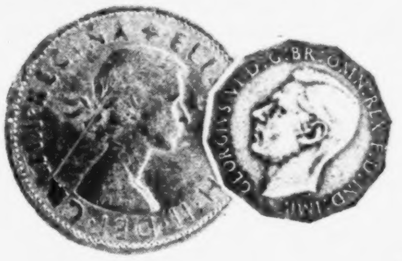

D

4 You have 4 shillings, 6 pence in your pocket and you spend half-a-crown. What is the name of the single coin you have left?

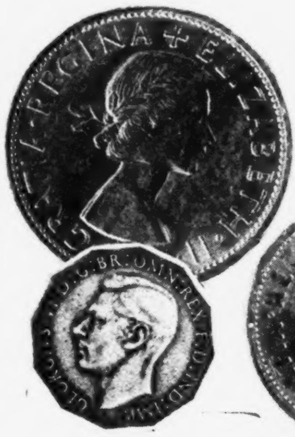
5 In words, write the value of a FLORIN.

6 You owe a friend the amount shown. As part payment you give the coins shown. Which coins do you still owe? A, B, C, or D?

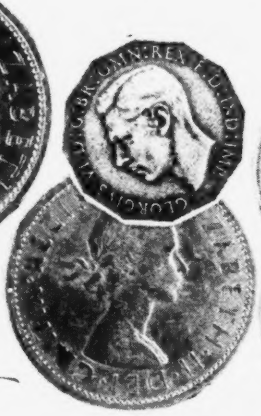
5/- −



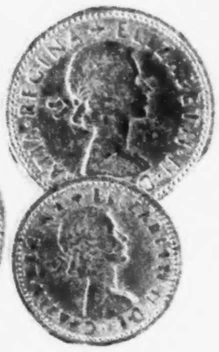
=



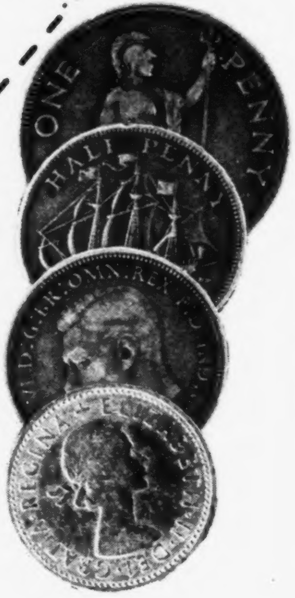
A



B



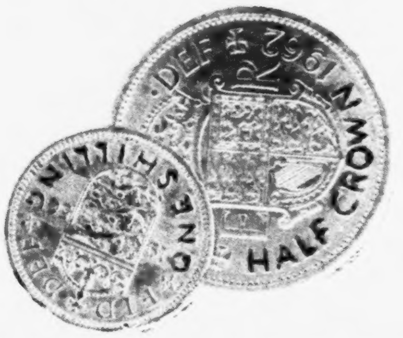
C


ONE PENNY
HALF PENNY

D

7 From the amount shown here, you spend the coins shown here. In words, write the amount you have left.

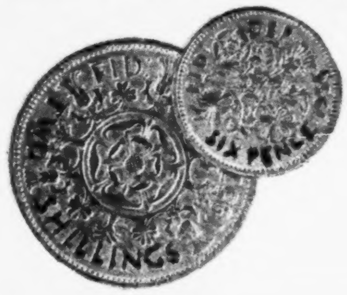
6/11 −


1962
HALF CROWN
ONE SHILLING

=

8 From the total amount shown here, you spend the coins shown here. In symbols, write the amount you have left.

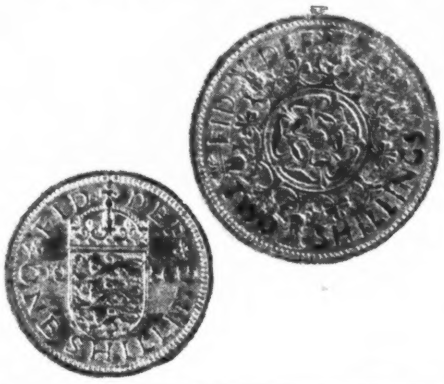
14 SHILLINGS, 9 PENCE −


SIX PENCE
SHILLINGS

=

9

With these coins you pay two-and-nine. The clerk hands you one coin in change. What is the name of that coin?



10

?

END OF TEST

I am a ____________________.
(boy or girl)

My birthdate is ____________ ________ ____________________
(Month) (day) (year you were born)

I thought this lesson was fun sort-of-OK boring
(circle one)

This is how you could make this lesson better:

ENGLISH MONEY PROGRAM

TRYOUT DATA

Test Item #	Correct Response	% of Sample Who Answer Correctly
1	C	.50
2	B	.43
3	C	.37
4	Florin,2s.	.27
5	2 Shillings, 2s., 24d.	.20
6	A	.20
7	3s. 5d., 3/5	.17
8	12s. 3d., 12/3	.17
9	3-Penny Bit, 3d.	.17
10	£2. 6d. 4½d.	.13

Average percent correct in sample group = 22.00+

ENGLISH MONEY PRACTICE WORKSHEET

Try to answer every question. Keep up with the filmstrip. Do NOT try to correct your mistakes. This is not a test.

SCRATCH AREA

A	18
1	19
2	20
3	21
4	22
5	23
6	24
7	25
8	26
9	27
10	28
11	29
12	30
13	31
14	32
15	33
16	34
17	35

DO NOT TURN PAGE UNTIL DIRECTED